The Relationship between Generative Grammar and a Speech Production Model

Kate Morton

Reproduced from the PhD Thesis *Speech Production and Synthesis*, June 1987.

'We are primarily concerned with establishing a model of speech production as the output of a generative *grammar.'* in: 'Some electromyography data towards a model of speech production' (1969) *Language and Speech* 12 (with Mark Tatham).

EARLY TRANSFORMATIONAL GENERATIVE GRAMMAR

In 1968 a generative grammar had three separate components: semantics, syntax, and phonology (Chomsky 1965). Phonetics was not considered part of transformational generative grammar (henceforth TGG) because language production below the level of phonology was not thought to be about cognitive processing (or, as was said at the time, mental processing). For the generative grammarian, phonetics was an autonomous subject describing the facts of speech. Phonetic models derived from acoustics, aerodynamics, mechanics, etc. For the TGG phonologist, phonetics provided a set of facts that could be used in phonological descriptions (Chomsky and Halle 1968). These descriptions had two functions:

1. to provide raw data about the sounds or articulations used phonologically in a particular language;
2. to provide a link to the external world by modelling how the output of phonology was realized as a speech waveform.

Some phonologists still consider this the subject of phonetics. They speak of 'the phonetic output', 'phonetic realization', or 'phonetic facts' in a general way, and regard phoneticians as describing the actual processes involved in speaking, or models of these processes. In fact, phonology considers speech processing to be automatic: its purpose is to realize as sound the requirements of the phonological, syntactic and semantic components which establish the linguistic properties of sentences.

Phonology was regarded as the component of the grammar which reinterpreted sentences by adding an abstract sound pattern to the output of the syntax. In modern terminology we could speak of phonology as being a re-encoding component providing an interface between the surface structure generated by syntax and the speech production mechanism which links to the external world. By analogy, a graphology would interface between syntax and the process of writing. At that time both speaking and writing were regarded as a-linguistic peripheral activities.

In contrast with earlier surface descriptive models, models of speech production were being proposed to account for the processes of realization underlying the surface soundwave (Fant 1960, Ohman 1964, 1966a, Kozhevnikov and Chistovich 1965, Ladefoged 1965, Kim 1966, Daniloff and Moll 1968, Fromkin 1968, MacNeilage 1970, Ohala 1970, Tatham 1970, Lehiste 1971). These models were concerned with describing automatic (that is, a-linguistic), physical processes such as neurophysiological , neuromuscular, aerodynamic and acoustic processes. They assumed an input from a higher level cognitive phonology, although they themselves were not describing cognitive processing. These models were relatively complex, and when formalized consisted of sets of rules.

MODEL BUILDING

Transformational generative grammarians made a distinction between the notions competence and performance (Chomsky 1957). The competence model characterized what a speaker/hearer knew of his language, and performance was about how that knowledge was used in converting a particular thought into a particular soundwave. Transformationalists were concerned only with modelling competence.

It is part of model building in general that the model need not describe the exact nature of the phenomenon. A useful characterization may consist of partial descriptions, or be simply a statement of general properties of the system. As much as is known (or relatively well established) makes up the larger part of the characterization, but for completeness and predictive power, some statements in the model can he hypotheses as to the nature of the phenomenon under observation.

Modelling in linguistics falls into this category: it is a set of established descriptions along with a number of hypotheses about language. linguistic models should not be regarded as derived entirely from empirically determined descriptions' see Action Theory and Description and Simulation): they cannot he directly related to objects and their behavior in the world.

Chomsky (1957) constructed a competence grammar that was intended to be descriptive (see Description and Simulation). The knowledge being characterized was said to be tacit; that is, not directly observable by the speaker/hearer. Techniques were developed by linguists to discover this knowledge, usually by questioning a native speaker/hearer about his intuitions concerning categories of meanings, grammaticality, and sound distinctions, by using test sentences. The linguist deduced what knowledge a speaker might draw on when he produces sentences. There were substantial differences among the phenomena being modelled, hence three components - semantics, syntax, phonology - were proposed. The data gathering techniques were similar for all three, and each was made up of descriptions and sets of hypotheses about its subject area.

COMPETENCE AND PERFORMANCE

At the time there seemed to be some misunderstanding about the relation between competence and performance. Some critics objected to the idea of this distinction, and some thought that performance was simply enacted competence. Competence characterized idealized language; errors observed in language use were felt to be a by-product of the performance process. Variability in speech was also considered an artefact of the performance mechanism.

Some confusion also resulted over the notions rule and rule-governed. In general, very few linguists believed that humans generate speech by running through a set of rules for each utterance, although a simulation could well do this (see Description and Simulation). It is now generally accepted that the rule format is just a formal way of describing a device that can generate an infinite number of novel sentences, and that the derivational trees generated by rules are a way of showing how apparent similar surface structures can be disambiguated, and also how dissimilar sentences can be related. Whatever the criticisms, TGG has been a fruitful model, characterizing knowledge that was needed for language processing: it has given rise to sets of hypotheses that have provided insights into the possible nature of language, and suggested hypotheses for researchers in other fields as to the nature of mind, and of processing within the brain.

A feature of language and speech processing which has proved difficult to deal with has been the apparent incompatibility between phonology and phonetics. One way of looking at the relationship between these components is in terms of the units that carry information from one to the other.

EXTRINSIC AND INTRINSIC ALLOPHONES

Standard TGG phonology (Halle 1979, Chomsky and Halle 1968) specifies an entry level to the phonology and an exit level from it. Between these levels there is a set of context sensitive

rewrite rules which map one level to the other. The entry level is called the level of systematic phonemics, and is also described as the underlying level. Units at this level are called units at the systematic phonemic level and also underlying forms. The exit level is called the systematic phonetic level, or the derived level; and units here are specified as units at the systematic phonetic level or derived forms. Thus

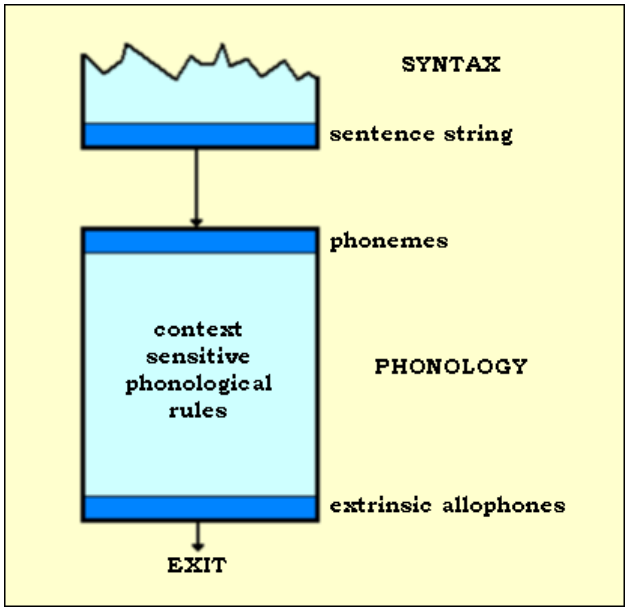

Fig. 1 Phonological levels.

At both levels, morphemes are characterized as strings of segments which may be represented in terms of a set of binary articulatory, acoustic or perceptual distinctive features. For example, a distinctive feature specification of the segment /a/ could in part look like this:

Consonantal	-
Vocalic	+
Diffuse	-
Compact	+
Grave	+
Flat	-
Voice	+
Continuant	+
Strident	-
Nasal	-

[from Hyman 1975, 351]

More traditionally this segment would be specified and classified among the other segments as:

[a] - unrounded open back vowel (Gimson 1984, p.329)

Segments at the systematic phonemic level are often (though strictly incorrectly, see Postal 1968) called phonemes, and they function in some ways similarly to the phoneme in traditional descriptions of speech sounds (Jones 1950, Abercrombie 1967, Gimson 1984). Segments at the systematic phonetic level are also sometimes called extrinsic allophones to

identify these as the result of applying phonological rules to the underlying phonemic string. Extrinsic allophones describe the result of cognitive processing: the result of choices by the language or dialect as to the realization of the underlying forms and which are under voluntary physical control.

Extrinsic allophones enter the phonetics component and are processed by the speech production mechanism. The output of phonetics is a speech waveform, which can be described by a set of intrinsic allophones. Allophones described at this point are the result of non-cognitive involuntary processes over which the speaker has no control (see Cognitive Phonetics). These processes are described in terms of neuromuscular control, aerodynamics, mechanics, and occur in the brain or at the periphery of the speech production system.

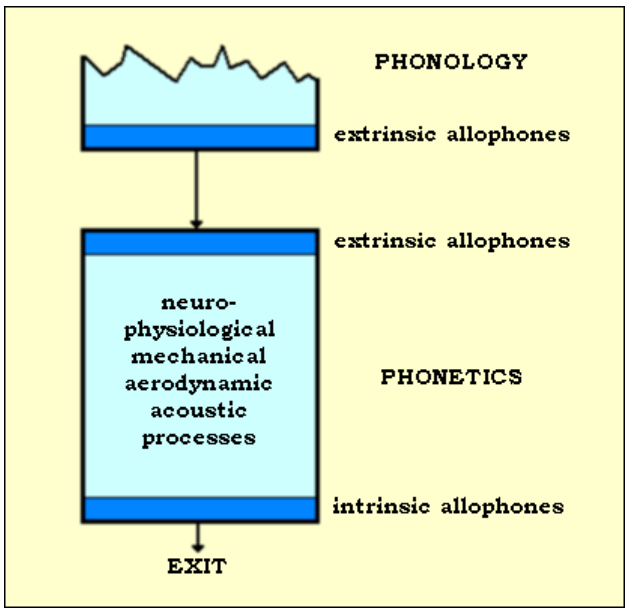

Fig. 2 Phonetic processes and levels.

These two terms, extrinsic and intrinsic allophones were originally defined by Wang and Fillmore (1961), and elaborated by Ladefoged (1967b) and Tatham (1969a). But they are used slightly differently now from when they were first proposed. Originally, they were labels assigned to variations of phonemes when they were realized by the phonetics.

However, the precise relationship was not discussed in detail; no distinction was made between these allophones in terms of their derivation, although it was implied that extrinsic properties occur before intrinsic properties. The notion of different levels, different classes at different levels and the idea that one is derived from the other came later.

For example: the English phoneme L (a capital letter is often used to symbolize an underlying segment, see Chomsky and Halle 1968) can be either velarized or palatalized since these features are not contrastive in the language. This segment undergoes a phonological process which maps it to two extrinsic allophones.

L → { lj / # - 'V } { lw / { - # } { - C } }

The phoneme L becomes the extrinsic allophone palatalized [lj] when it occurs word initially preceding a stressed vowel, and becomes velarized [lw] when it occurs word finally or preceding a consonant.

Phonological mapping is conditional on how a segment stands in context with other phonological segments (including boundaries). The position of L within the string forms the condition on the rule. There are no physical conditions, such as co-articulation, mechanics of articulation etc. , on the rule. For example, the following rules are possible in principle; they

do not violate physical conditions, and could occur in standard southern British English although they do not:

L → { lj / - # } { lw / # - 'V }

L → lj / in all contexts

Going back to the first set of rules which derive extrinsic allophones for southern British English from the underlying phonological unit L, the extrinsic allophones [lj] and [lw] now undergo coarticulatory (physical) processes in the phonetics component.

lj → lj [+ front] / # - V [+ front] ... → lj [+ back] / # - V [+ back]

lw → lw [+ front] / # - V [+ front] ... → lw [+ back] / # - V [+ back]

These rules state that some fronting and retraction are applied to both palatal [lj] and velar [lw] in correlation with the degree of fronting and backing of the adjacent vowel. In each case between the two extreme of [+front] and [+back] there is the possibility of rules deriving [lj] or [lw] with different degrees of fronting or retraction depending on the frontness and backness of the vowel context. (Notation conventions vary. For symmetry, these rules are expressed in a notation like that usually used in phonology.

This produces sequences of intrinsic allophones derived from extrinsic allophones, and extrinsic allophones derived from phonemes - not extrinsic and intrinsic allophones both derived equally from phonemes, which seems the case in Wang and Fillmore (1961) and Ladefoged (1967b). Subsequently Ladefoged recognized this point (1971), and now uses the term coarticulatory allophone to refer to an allophone derived by coarticulatory processes (1982).

However, the status of intrinsic allophones is not quite clear. The difficulty arises because of the implication that lj [+ front] and lj [+ back] are 'real'. There is general agreement that phonemes and extrinsic allophones are abstract symbolic representations. But there seems to be a divided view about whether intrinsic allophones are real or symbolic. This may be the result of a model which describes data in terms of levels of representation: is the label intrinsic an abstraction directly based on data from the physical non-cognitive world, or a symbolic device on the same level of abstraction and relating to the same kind of object as, for example, extrinsic allophones? Is an intrinsic allophone an object or a symbolic representation of an object (see Description and Simulation)?

However arrived at , the concept of language-dependent extrinsic allophones (which describe the intention of the speaker to pronounce sounds in a particular way) and the intrinsic label attached to the output of the physical production process have enabled a clearer picture to be formed of the relationship between phonetics and phonology.

To summarize: from a phonetic perspective and without using the Chomsky-Halle terms, we can recognize three levels of symbolic representation in speech production. The levels are referred to in terms of the type of symbol used to characterize strings. Thus, at the beginning of the phonology the characterization is in terms of phonemes - minimal units whose main function is to identify and distinguish between morphemes. After the application of context sensitive phonological mappings strings are represented in terms of extrinsic allophones - units whose main function is to characterize how the speaker wishes utterances to be pronounced. Mappings in the phonetics derive objects represented symbolically in terms of intrinsic allophones - units whose main function is to characterize the final phonetic output of the speech production process.

GENERATIVE GRAMMAR AND SPEECH PRODUCTION

The idea expressed in the quotation at the beginning of this article is therefore too simplistic. Speech production cannot be the output of a generative grammar because it is modelled differently from the rest of the grammar. At best it could be correlated with phonology. That is an event in one description might be linked with an event in another. Pairs of objects in each domain might be relatable but not necessarily in a cause and effect relationship or in the

sense that the output of one constitutes the input to another. The model generally accepted in the late 60s is shown in outline in Fig. 3.

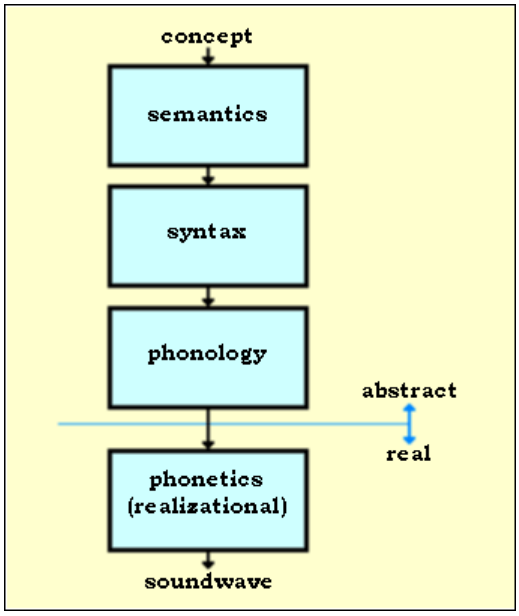

Fig. 3 Late 60s model relating generative grammar and speech production.

INCORPORATING PHONETICS INTO LINGUISTICS

During the 60s , linguistics developed from a surface description of language into a model which described language as a set of underlying structures related to the surface description by a set of rules. However, phonetics, which describes speech, was not integrated into the new approach. But since phonetics describes how the surface sentence description can be mapped to the external world, it is part of language processing and the relationship should be expressible in the same general grammar model. Thus, if the linguistics model is concerned with characterizing knowledge bases, then a phonetics component forming part of the grammar should also be treated as a knowledge base with the same type of facts and rules as other components. This principle of organization would make phonetics compatible with the grammar and especially with phonology, its closest related component.

Traditional phonetic studies about speech production do not distinguish between competence and performance. But if we are to apply the generative approach of linguistics to phonetics, this distinction is necessary. (The benefit to phonetics of making the distinction is discussed in Tatham 1980 and Repp 1986.)

Fig. 4 illustrates an early 70s model which adds a competence phonetics to a transformational generative model of language.

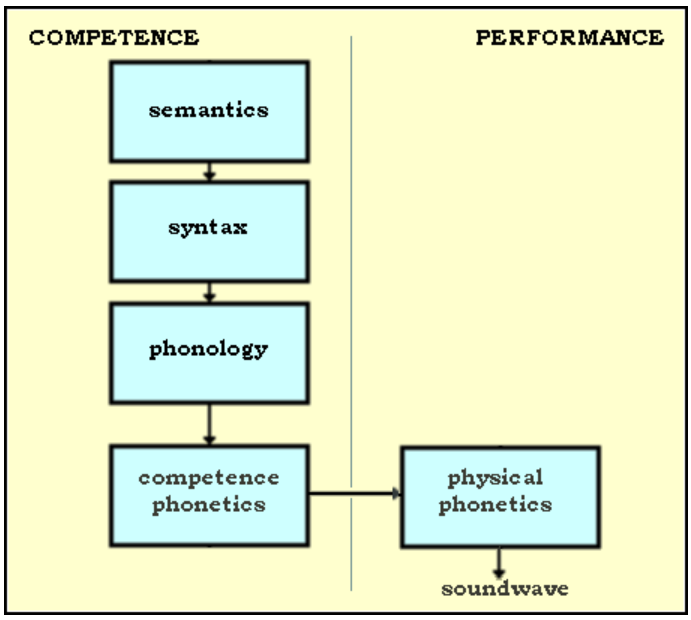

Fig. 4 A model of language production distinguishing between a physical phonetics and a competence phonetics.

If a competence component were to characterize phonetics knowledge, two questions arise:

- what is the kind of knowledge this component could describe, and
- what in a description of a composite competence/performance model can be assigned to either a competence or a performance phonetics.

When a competence phonetics has been determined, we then need to ask

what are the facts and relationships between these facts, and

how might these facts used by the speaker.

Such an approach distinguishes between what we know about speech and what we know about how to speak, i.e. competence and performance. Phonetics will still be unique, however, since it is concerned with describing actual performance.

Describing phonetic performance necessitates taking decisions as to the best method for representing data contained within the knowledge bases. For example, information about producing fast speech could be represented as a distinct mode of speaking, necessitating its own database and rules. Rules would call the alternate mode when fast speech is appropriate. Alternatively it could be represented as a departure from some normal mode.

THE CURRENT VIEW IN LINGUISTICS

During the last decade an interest in knowledge representation has developed in many areas. In linguistics/phonetics it can be useful to regard competence as the knowledge base which characterizes facts , and rules which relate these facts about the knowledge a native speaker has of his language. The human being's knowledge base is represented in the linguistic model as a static description which is called upon, or accessed, by some other device which encodes knowledge from which a sentence is generated. The knowledge base is consulted to

- ascertain how a specific sentence might be generated,
- determine how different sentences can be related descriptively, and
- determine the underlying structure of a sentence or class of sentences.

This type of model can provide the underpinning for simulation (see SYNthEX and Description and Simulation).

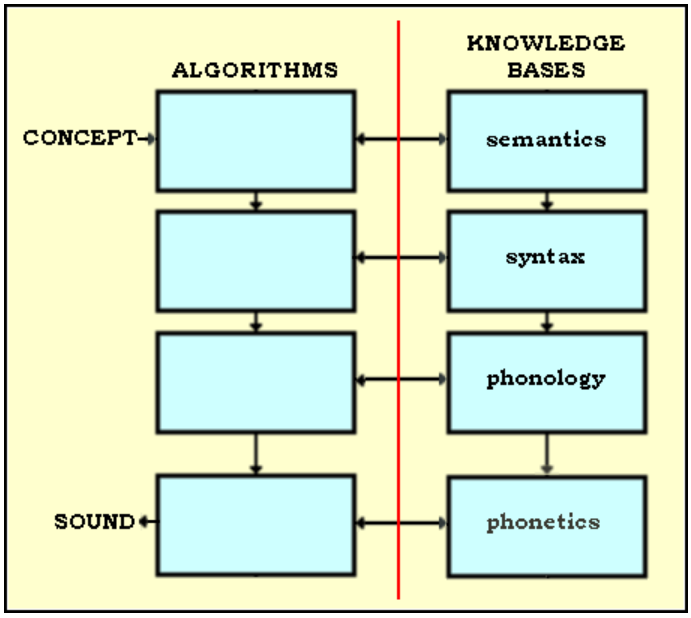

Fig. 5 Encoding concept to soundwave showing the algorithm's access points to linguistic knowledge bases.

The model illustrated in Fig. 5 shows encoding from concept to soundwave by a series of algorithms. Each set of algorithms makes decisions based primarily on information contained within its associated knowledge base. The overall encoding process consists of four stages. At each stage the algorithm selects rules from the knowledge base as input to the encoding process Each component knowledge base is a list of the facts about the knowledge a speaker has, and rules specifying all possible relationships between facts for each level. For each particular utterance only a subset of possible facts and relationships is drawn out by the encoding algorithm. The selection procedure is governed by the specific requirements of the input device - the concept generator - on any one occasion. (Although the Fig. presents the relationship linearly, the procedure is not necessarily sequential.

This model adds a phonetic component to the linguistic grammar similar to the other components in the way it functions and in its characterization of knowledge about the system. In phonetics, two kinds of facts and rules are represented in the knowledge base:

- knowledge about the nature of the physical system and how it works,
- knowledge about the physical system upon which the decision processes draw: facts about anatomy, neurophysiology, aerodynamics, acoustics, etc.

Thus a coherent relationship between linguistics and phonetics is established without a break in type of theory and including data which previously could be found only within traditional autonomous phonetics.

The model shows the distinction between competence and performance at the phonetic level: knowledge bases as the competence model of this part of the system, and the decision algorithm as performance. Before the model is run, the decision making processes are also characterizations of potential: that is, all decisions within the constraints found in the knowledge bases are possible. But a single run of these procedures, triggered by an input from the concept generator, will be a performance event. Which decision procedures are selected, and, in turn, which data and rules are selected from the knowledge bases are determined by the requirements of this particular encoding job, but the description of the entire system is constant as a competence characterization of the system's total potential for speech production.

One function of the decision processors is to consult the knowledge bases with the question: is this event possible? If the answer is 'yes' the system can be manipulated within

the constraints specified in the description If the answer is 'no' , what is proposed is physically impossible.

The content of the knowledge bases 15 derived from performance events. Thus phonetics, through experimentation, initially models performance data by describing

- the physical parameters involved in phonetic realization of extrinsic allophones,
- the manipulation of these parameters to obtain versions of these allophones under varying conditions of speaking, and
- the predicted acoustic output.

This model , adding a phonetic component according to the practice of transformational generative grammar, can be elaborated to take into account the importance of distinguishing between a descriptive model and a simulation based on that model (see Description and Simulation).

SUMMARY

Chomsky (1957) changed the direction of development of language description by introducing transformational generative grammar into linguistics. Linguistic theory moved toward a description of language which is about the general principles which might underlie surface utterances. This approach has tried successfully to show that the process of generating language can be described by a competence model using a relatively small set of rules. Many phoneticians were also interested In accounting for surface phonetic data by postulating a general underlying system. This can be seen in studies of coarticulation which was viewed as rule-governed target missing (Ohman 1964, 1966, Kim 1966, MacNeilage and DeClerk 1967).

We were interested in the 60s in extending the ideas of explicitness and rigorous formulation that were already a feature of phonetic research during the 50s and 60s. However, this earlier work was essentially data based, rather than particularly theoretically motivated. That is there was no well formulated theory of phonetics to underpin the experimental work. Most of the experiments of the time observed data and described details of speech, attempting to show how the observations and results might fit into an overall model explicitly designed to explain speech production processes. A close relationship with the rest of the language production process was not formulated clearly at that time although descriptions about both language and speech processing were being sought.

Transformational generative grammar modelled the components of the grammar as expansions of underlying forms. The general process of successive expansions is known as the Translation Model (Fowler et al 1980), and is discussed in the next section.

References

References are all to items listed at the end of the Thesis.