Journal of Phonetics (1982) **10**, 37–72

The concept of phonological form and the activities of speech production and speech perception

Per Linell

Department of Linguistics, Uppsala University, Box 513, S-751 20 Uppsala, Sweden

Received 17*th February* 1981

> "On the one hand phonology, which stops at phonemes and their distinctive components, and on the other hand speech analysis, which discards all questions of purpose, seem equally obsolete in their narrow isolationism and futile abstractionism."
>
> (Jakobson & Waugh, 1979, p. 48)

Abstract: This paper deals with the relation of phonology to the phonetic theory of speech performance. In particular, it is concerned with the proper interpretation of the concept of phonological form.

The paper is organized in two parts. Part I contains a discussion of the main requirements that can reasonably be placed upon the phonological form as an explanatory construct in phonetic theory. It is proposed that the phonological form of a given linguistic expression be construed as a *phonetic plan*, i.e. a plan for performing a phonetic-behavioral act that can be perceived and comprehended by language users as the linguistic expression in question.

In Part II it is discussed how, more exactly, such a phonetic plan should be understood in relation to the activities of speech production, speech perception and phonetic imagery. Some aspects of an action-theoretical approach to phonetic behavior are also explored and contrasted with the prevailing so-called translation theories of speech production and perception.

Part I

1. Introduction

Throughout the history of modern linguistics the sound structures of languages have been studied by two rather different disciplines, phonology and phonetics. This means that we have lived with a perpetual problem of defining the proper relation between these two fields. What is the true nature of the subject matter of phonology? How can phonology be assigned a function in an explanatory phonetic theory of speech production and/or speech perception? Or should we rather assume phonology and phonetics to be mutually irrelevant? Such questions have been answered in different ways in different epoques of the history of the phonetic sciences. Rather often the relation of phonology to phonetics has been conceived of as quite distant and indirect. In my opinion, this is an embarassing fact.

It is customary to discuss phonology and phonetics in terms of *form* vs *substance* (cf.

0095-4470/82/010037+36 $02.00/0 © 1982 Academic Press Inc. (London) Ltd.

Fischer-Jørgensen, 1975). Within modern linguistic theory these notions are derived from Saussure and further developed by, among others, Trubetzkoy (*Sprachgebilde* vs *Sprechakt*) and Hjelmslev. Here, I will approach the phonology–phonetics issue by focusing on the notion of phonological form. However, I will *not* consider a Saussurian notion of form in the sense of the entire linguistic structure underlying and/or inherent in phonetic behavior (substance). Rather, I will discuss a narrower (though not unrelated) concept, i.e. the notion of phonological form (or phonological representation) as it is used in contexts like "the phonological form, or representation, of a certain expression, e.g. the word (phrase, sentence, utterance etc.) X". Thus, we often talk about the phonological form of, say, the expression *They want you* or the phonological representation of the word *extraordinary*. What is the meaning, if any, of the expression "the phonological form of *X*"? What kind of reality may such a term refer or allude to, and what role does this "reality" possibly play in speech performance (phonetic behavior)?

In my subsequent discussion I will assume that it is indeed possible to find a reasonable empirical interpretation of the concept of "phonological form" (or "representation"). However, this does of course *not* presuppose any acceptance of the ways in which these terms and the associated concepts have been used within generative phonology. On the contrary, I would argue that the abstract morpheme-invariant (morphophonemic, "systematic phonemic") representations that generativists call "underlying phonological representations" cannot be assigned any reasonable or fruitful, empirical interpretation (Linell, 1979). On the other hand, my discussion will have a great deal in common with a view of phonology that I take to be typical of many "concrete-minded" phonologists, e.g. many representatives of classical structuralism (particularly in Europe) and also the more recent "natural phonology" (Donegan & Stampe, 1979; Stampe, 1979; Wojcik, 1979).[1] Also I will refer to several recent proposals in phonetic theory.

2. Some requirements on the concept of phonological form

2.1. The utterance (word, phrase, etc.) *X* referred to in "the phonological form of *X*" is a string of spoken language, a phonetic signal and not some other, e.g. graphic, manifestation of language. This is naturally important and will have an impact on how we conceive of the phonological form. To be more precise, *X* is a stream of phonetic *behavior*, a series of (maybe partially overlapping) *meaningful phonetic events* distributed in time. Phonology specifically deals with the outer physical form, the "*sounds*", of this behavior, not with its grammatical structure.

Obviously, there are several ways in which the implications of this condition can by bypassed. Thus, the behavioral time-distributed nature of the phonetic signal is often forgotten by phonologists who deal with phonological forms as if they were indeed just strings of graphic symbols (as in our notations on paper).

An even more radical way of abstracting away from the phonetic basis is the glossematic fashion of treating phonological forms as structures of oppositions with no phonetic content at all. Yet another unfortunate proposal amounts to treating (most of) morphology and morphophonology together with phonology proper in one monolithic phonological component as in (orthodox) generative phonology (e.g. Chomsky & Halle, 1968).

2.2. The second condition involves the so-called type-token problem. One may note that expressions like "the expression *They want you*" or "the word *extraordinary*" are ambigu-

[1] Cf. also the so-called natural generative phonology (Hooper, 1976).

ous. Do we mean the expression *They want you* taken as a linguistic type, i.e. the abstract entity belonging to the set of linguistically different expression units that can be constructed given the system of the English language, or do we mean one particular utterance token, the expression *They want you* uttered by a particular speaker in a specific situation at a particular time? This is an important question which is usually ignored in phonological theory where it seems to be tacitly assumed that the expression *type* has one unique phonological representation which also underlies (is present in, or manifested by, or realised through) all the particular tokens ("occurrences") of the expression in various utterances. If so, the phonological form of the type is also the phonological form of the tokens.

Accordingly, it may be argued that although the specific utterance tokens vary in different phonetic details, they are still classified (by linguists and language users) as *tokens of certain linguistic types*. Thus, *They want you* may be pronounced in many different ways, e.g. with or without internal pauses, the syllabic nucleus of *want* may be more or less nasalized, the /n/ may be more or less absorbed into the nucleus, the /t/ may be clearly aspirated or it may join with /j/ of *you* to form [ʧ], the vowel of *you* may vary from [u:w] to [ə] etc. In fact, *They want you* may vary in its pronunciations at least from [ðeɪwɒnt^{h} ju:w] to [ðewõʧə], and *extraordinary* from, say, [ˌɛkstrə'ɔ:dɪnərɪ] to ['strɔnrɪ] (cf. Brown, 1977). Still, we would say that in some important respect all tokens are the same, something which seems to be a social, psychological and linguistic reality. As long as they are pronounced in a reasonably proper way by speakers of the same dialect, they count as instances of one and the same linguistic unit, and speakers and listeners typically believe that they are the same (perhaps even phonetically), thus often underrating the phonetic variation that actually occurs. It is often thought that the phonological form pertains to the invariant features that define the expression type in the language, thus distinguishing it from all those other expressions (expression types) in the language which are phonologically-phonetically different. Sometimes the phonological form is said to be *equal to what is invariant* and common to the tokens.

The invariance condition is intuitively well-motivated but far from unproblematic. For one thing, it must not be interpreted so as to imply that a certain set of phonological-phonetic features (those present in the phonological form) must be physically present in all tokens; many tokens are strongly "reduced" and have "lost" many of the features that are present in careful pronunciations. I will argue below (Section 6) for the traditional assumption that the phonological form is based upon the phonetics of a careful pronunciation.

However, there are more problems with assuming the existence of just one unique phonological form underlying all utterance tokens of what counts linguistically as the same expression. Clearly, one must allow for cases with several distinct forms (*either* /aiðə/ – /i:ðə/). Apart from such clear cases, there are countless cases in which speakers with partly different dialects pronounce the same words consistently in more or less distinct ways. To what extent can we say that such speakers share the same language with the same phonological forms? The answer to this question will probably depend on the terms in which one chooses to define language (cf. Section 2.5). Within the perspective to be proposed here we must allow for the possibility that speakers in a heterogeneous speech community have partly different phonological forms of the same words.

2.3 Thirdly, the condition of *linguistic relevance*, which is closely related but not equivalent to the invariance condition, is the corner-stone of classical phonological theory; the phonological form of an item comprises all and only those phonetic features that define the linguistic identity of this item and distinguish it from all other possible items in the language. In slightly different words, the phonological form of an utterance comprises all the phonetic

information that is "linguistically relevant" in that it prevents the utterance in question (and its meaning) from being confused with any other conceivable utterance (with a different meaning). Thus, according to the tradition the phonological form contains only "distinctive" features, whereas "predictable", "allophonic" features are absent.

Though the gist of this reasoning must no doubt be reflected in any adequate theory of phonology, the condition under consideration is not free from problems either. One aspect of the problems involved concerns the absolute distinction between "distinctive" and "non-distinctive" ("redundant") features that the strictly phonemic theory introduces. There is reason to suspect that psychologically there is no such absolute and clearcut distinction.

Furthermore, when "linguistic relevance" is defined in terms of capability of distinguishing meanings, we do not consider all kinds of meanings that are commonly conveyed in normal verbal communication. Traditional linguistics has been preoccupied with cognitive meaning, leaving emotive meaning aside, and this fact is reflected in phonological theory. The way an utterance is pronounced (there are variations in tempo, loudness, pitch span, direction of pitch, voice quality, articulatory precision, pausing, etc. cf. Brown, 1977, chapter 7) conveys many attitudes, emotions, physiological and mental states, etc. Yet these differences would be handled by allophonic rules rather than being reflected in the phonological forms as the latter entities are usually conceived of. A reasonable conclusion would then be that we may work with phonological forms roughly according to the tradition but that phonology cannot be exclusively concerned with those forms. The study of phonological, in particular allophonic, rules and their functions is also necessary. While the features specific in phonological forms are functionally basic and presumably psychologically central, the predictable features derived by rules are not linguistically irrelevant; they have important cosignalling functions.[2]

2.4. Fourthly, an adequate theory of phonology must fit into a coherent and plausible theory of language and human communication in general. As regards a model of linguistic structure or linguistic competence, it is important to *delimit phonology properly* from other components, in particular morphology and lexicon. If phonological regularities are confused with morphological, morphophonological or lexical regularities, one may arrive at theoretical constructs like abstract morpheme-invariant forms (as in e.g. Chomsky & Halle, 1968), which cannot be incorporated into a theory of linguistic behavior (Derwing, 1973, Linell, 1979).[3]

2.5. These conditions anchor phonological forms in something outside and beyond the utterance tokens themselves, namely in the linguistic system of expression types. However, this still leaves the question open as to how the linguistic system itself should be understood. What kind of existence has language? What kind of ontological status have phonological forms? If we want to assign an explanatory function to phonological forms in a theory of phonetic behavior, the answer seems obvious; phonological forms must be psychologically real. In some way or another they must be understood as aspects of the speaker-listener's language-specific phonetic competence. A (spoken) language is a complex system of dispositions to experience and structure, produce and understand phonetic behavior. These dispositions are socially shared to a large but not necessarily total extent. The various behavioral acts are conventionally appropriate to social situations and represent attempts

[2] On the notion of cosignalling, cf. Korhonen (1969), Dressler (1977), who mainly apply it to morphophonology.

[3] The proper delimitation of phonology relative to other components of linguistic competence is an issue which will not be further discussed in this paper. See Linell (1979).

to communicate meanings of various kinds. Thus, phonological forms are socially shared rather than idiosyncratic individual facts.

All this calls for a treatment that can do justice both to *psychological and social reality*. Unfortunately, linguists have often been too onesidedly biased in their metatheoretical preferences. On the one hand, a psychological interpretation of phonological forms, and of language in general, is far from universally accepted. In fact, the dominant attitude among classical structuralists seems to amount to regarding phonological forms, and language in general, either as pure abstractions or as social values (irreducible social facts in the Durkheimian vein). On the other hand, generativists and information processing theoreticians relying on certain traditions in American individual psychology have developed models of phonology and grammar in which the social aspect is almost completely lost. Instead of regarding rules as conditions on publicly observable phonetic behavior, these models are centered around behaviorally uninterpretable concepts like abstract morphophonemic forms, long derivations involving opaque and extrinsically ordered rules conceived of as deep-seated mental processes, etc. Various critics have rightly argued that such a theory makes the social origin and social sharedness of rules inexplicable. Our general theories of linguistic competence and performance must possess ecological validity in a world of human language users acting in normal social contexts, i.e. not in a world of idealized speaker-hearers in absolutely homogeneous speech communities.

3. Some assumptions about language use: construction and execution

In approaching a reasonable and more precise interpretation of the concept of phonological form, I will now state some general assumptions about linguistic structure and communication that will provide a suitable framework for our discussion.[4]

The basic function of language is communication between human beings in different types of social situations. Language and its structure are shaped by the conditions on communicative situations and can be properly understood only in this kind of perspective.

In a typical communicative situation, two persons, A and B, interact. In one communicative turn, A wants B to know, feel or do something particular, and this may be said to be his intended message M. M can be communicated by verbal and/or non-verbal means. If we consider only speech communication, what A does is to produce one (or several) utterance(s).

The verbal act of using an utterance for conveying a certain meaning can be regarded as a complex means–end hierarchy of acts and subacts. The act of producing a certain utterance with a certain conventional meaning is itself a means for achieving the goal of communicating the intended meaning. We will assume that the act of producing an utterance has two aspects, *construction* and *execution*. Since most utterances do not exist as ready-made entities in the mind of the speaker he must construct them by applying certain grammatical operations (subroutines, construction methods) to those lexical building-blocks to which he *has* access. What the speaker constructs may be thought of as a plan for producing a certain string of phonetic behavior. What distinguishes meaningful phonetic behavior from "mere vocalizations" is that it is subject to certain grammatical and phonological (phonetic) conditions, and these conditions must be specified in the speaker's utterance plan. For the utterance to be observable by the listener, it must of course also be executed, i.e. pronounced. The concepts of construction and execution will be further discussed and clarified later on (Section 7).

[4] For much more comprehensive discussions, see Allwood (1976).

4. Expression plans and linguistic types

An adequate phonetic theory of speech production and perception must be, at least partly, based on action-theoretical concepts. Processes "at the phonetic level" are embedded in acts of communication and therefore subordinated to communicative intentions. Thus, we argue that the production of an utterance involves the construction and execution of an *expression plan*. Such a plan specifies the *linguistic identity* of the utterance, i.e. it tells what words, grammatical constructions and prosodic patterns are intended and produced. For example, every time a speaker intends to say *They want you*, he constructs the same expression plan specified by the same invariant semantic, grammatical and phonetic properties. In other words the concept of expression plan accounts for the *functional equivalence* of the various physical manifestations. The plan (intention, goal) is what is constant, while manifestations vary.

The expression plan of an utterance can be analyzed in terms of its grammatical (syntactic, morphological) aspects and its phonetic aspects. The basic units are word forms and prosodic patterns. When we (here) talk about *phonetic plans*, we are only thinking of the phonetic aspects of word forms, which is what corresponds to the traditional notion of phonological form. Thus, an expression plan specifies the outer form of a given utterance, and it can be sliced up in two dimensions yielding four subparts, only one of which corresponds to phonological forms.

(1) *Expression plan*

Components / Aspects	Word forms	Prosodic patterns (phrase and sentence prosody)
Grammatical aspects		
Phonetic aspects	Phonological forms = phonetic plans (in a restricted sense)	

Our notions of expression plan and phonetic plan are both linguistically and psychologically motivated. On the one hand, the expression plan states the *linguistic equivalence* (or identity) of the various phonetic tokens, or, if you will, it expresses the linguistic-phonological *meaning* (content, import) of these manifestations of the same expression type (word form, phrase, sentence, etc.). On the other hand, it represents a direct psychological reality in that it constitutes a *plan* for (real) phonetic behavior, a plan used (constructed or activated) by speakers and listeners.[5] Speakers intend to produce the same kind of phonetic signal every time they are about to pronounce a certain expression. (Actually, this is not quite true, since speakers may intend to pronounce the various tokens in different ways, and therefore the execution phase must be assumed to involve some additional planning, primarily having to do with speaking style, cf. Section 7). Furthermore, when speakers consciously imagine the pronunciation, they tend to produce a careful, quasi-invariant manifestation, which seems to mirror the full structure of the phonetic plan rather closely (see Section 5). Listeners, on their part, reinterpret the same phonological form into the distinct phonetic occurrences and typically believe that the latter are much less phonetically dissimilar than

[5] Thus, note the logical connection between the meaning (import) of the behavioral act and the intention or plan behind it.

they actually are. Thus, the notions of expression plan and phonetic plan are necessary for the explanation of both speakers' behavior and listeners' behavior. In addition to this, we must of course explain the simple observational fact that, despite certain physical differences, the various occurrences of the same expression do share many phonetic properties.

Linguistic units, expression types, may be conceived of as dispositions to behave phonetically in particular ways. Phonetic behavior and its underlying plans may be characterized in terms of either articulatory gestures or the resulting acoustic events. Thus, in a specific situation, the same particular expression plan may be seen as *a plan to make certain articulatory gestures* such that the sequence of behavior fulfil certain phonological and grammatical conditions defined by the plan, *or* as *a plan to produce* patterns in the acoustic field such that a perceiver may interpret them as *a certain sound event* having the properties specified in the plan.[6] The phonologically relevant features of a certain word (say *day, they, want, you, cats, extraordinary*) (whatever these are) may thus be characterized in both articulatory and acoustic/perceptual terms. However, ultimately features must be assumed to be perceptually determined (Section 16). Thus, the articulatory gestures are the way they are because they are intended to produce certain acoustic patterns which in their turn must be amenable to perceptual identification by listeners in just the ways required.

By way of summary, I propose to *define phonology* as the study of *language-specific aspects in the phonetic behavior* (pertaining to different languages), or, alternatively, as (the study of) the rules for the conventional use of phonetic mechanisms in specific languages. *The phonological form of a certain linguistic item in a given language may be seen as a plan* (disposition) *to produce a certain phonetic event meeting those phonetic conditions that define the linguistic identity of the item* (make it different from other possible items). This will prove to be a fruitful interpretation of the concept under discussion. Thus, it copes with all the requirements discussed in Section 2, and it ties in very well with recent phonetic theories of speech production, which have strongly emphasized the role of target specification and motor equivalence (defined by MacNeilage (1970) as the "achievement of relatively invariant motor goals from varying origins"). In a similar vein, other scholars have pointed out that speech is "goal-oriented" (Chistovich & Kozhevnikov, 1969) or "output-oriented" (Lindblom, 1979) (cf. also Section 14).

In Part II of this paper I will modify my characterization in some minor but important respects, so that the concept will be assigned a proper interpretation both in the theory of speech production and in the theory of speech perception.

5. Phonological forms as plans for phonetic behavior

Thus, we have come to the conclusion that a phonological form is a plan to perform a certain type of phonetic act, i.e. to produce a sound signal with certain specific phonetic properties. We can further assume that such plans are associated with "concrete" forms, i.e. word forms[7] (or possibly stems or morphs, i.e. *parts* of word forms) and some phrases.

[6] Most of the concepts applied in these definitions are discussed by Öhman (1975*a, b*).

[7] The relevant domain of a phonetic plan is, I believe, the word. However, "word" is a vague term which may be specified in different ways ("word" in the senses of "lexical item", "morphological word (= word form)" and "phonological word"). I assume that "morphological word" (= word form) is the most likely candidate for constituting the domain of a phonetic plan. However, the other concepts may also be relevant. Possibly, one has to work with different kinds of domains in typologically different languages. For some discussion, see Linell (1979, chapter 8). In Linell (ibid. chapter 12) I show that morphemes cannot be assumed to be linguistic primes.

Thus, each word form has (at least) one phonetic plan that specifies its phonological identity. When a speaker intends to pronounce the word form in question, he aims at realizing those phonetic effects that are specified in the plan. I disregard for the moment the specifications due to the choice of speaking style (see Section 7).

Naturally, the properties specified in a phonetic plan are *real* phonetic properties that can be physically manifested[8] when the word form is actually pronounced (although not all properties are always manifested, since speakers often execute their plans only partially). Properties specified in phonetic plans pertain directly to real phonetic events that speakers attempt to produce, not to some mystical abstract entities as in many generative-phonological "underlying forms". It would be nonsense to assume that speakers plan something which can never be executed. In addition, one should recall, in connection with this, that linguistic rules or conditions must be learnable, which means that they must be based on facts that are socially accessible and observable. Therefore, this theory cuts the Gordian knot of the so-called abstractness problem in contemporary phonological theory in quite a radical – and highly desirable – way.

The assumption that the tokens of a certain expression type (word form) have one common phonological form amounts to saying that only certain properties of phonetic events (sound signals) are of primary linguistic relevance. These properties define the linguistic status of all the tokens (their linguistic equivalence). The properties specified in the phonological form (phonetic plan) can also be regarded as the *item-specific rules* or conditions that the tokens must satisfy (Linell, 1980); thus, the conditions specified in the phonological forms of *they* or *want* are rules for the pronunciation of these particular items.

We may assume that the specifications given in a phonetic plan distinguishes the word in question from other possible non-homophonous words. It is of course an empirical question to find out exactly what these linguistically relevant properties are in the various individual cases and whether they may be characterized in terms of general principles. However evidence of many types (e.g. judgements of identity, similarity and correctness of phonetic strings, reinterpretation of deviant sounds, symmetry in phonetic space, adaptation to secondary dialects, slips of the tongue, child language, nature of morphological operations) indicates that surface phonemic contrasts are relevant for determining them (e.g. Linell, 1979 chapter 5).[9]

6. The role of careful pronunciations

Words may be pronounced in many ways, but their underlying phonetic plans pertain to the most *careful pronunciations*. By "most careful pronunciation" I mean the result of a competent speaker's effort to pronounce the word in question as carefully as possible in order to manifest all the phonetic properties that keep the word phonetically distinct from all other (perhaps phonetically very similar) words in the language. Note that such a "careful pronunciation" is something of an "ideal form". Thus, it cannot in general be equated with the so-called citation form (pronunciation in isolation), since citation often involves neutral-

[8] This does not exclude the possibility that some are "abstract features" in the sense of "cover features" (Ladefoged, 1971). I take a cover feature to be the emergent, complex property built upon specific configurations of values along several phonetic dimensions. Many phonological features are of this type (e.g. stress and other prosodic features).

[9] This view of phonology is not committed to the assumption that phonemes are the basic units. Although surface-phonemic contrasts are no doubt important, an extreme structuralist view of phonemes simply as redundancy-free phones must be rejected (e.g. Bailey & Maroldt, 1979).

izations of distinctions present in careful pronunciation. Citation often involves *hyperarticulation* ("sharpening" or "elaboration" acc to Linell, 1979). Consider, e.g. the French words *nier* "deny" and *briller* "shine"; clearly, these words are phonetically distinct in that *nier* ends in /iˈe/ and *briller* in /iˈje/. This distinction is brought out in careful pronunciation, and the words are also reduced in fast speech in different ways. However, in elaborated pronunciation *nier* becomes [ni · ˈje ·], i.e. /i/ is lengthened and a hiatus-breaking [j] is inserted, thus making the word rhyme with *briller*.

Thus, actual pronunciations may *deviate* from the ideal, careful pronunciation *in two directions*, i.e. *hyperarticulation* (elaboration) in formal, stylized speech, citation, etc. and *hypoarticulation* (reduction) in casual, sloppy speech. The former involves *fortition* processes, the latter *lenition* processes (cf. Donegan & Stampe, 1979). Wojcik (1979) notes the following:

> Two types of processes exist to accomodate two somewhat contradictory needs in speech production. Fast and sloppy speech is characterized by processes which simplify articulation. Assimilation, coalescence, reduction, and deletion carry out simplification. Donegan & Stampe (1978, 1979) have called such processes *lenitions*. Slow and careful speech is characterized by processes which enhance the acoustic properties of segments. Dissimilation, polarization, strengthening, and deletion increase phonetic salience. These are called fortitions.
>
> The two types of processes usually have contradictory effects. For example, my casual pronunciation of *parade* and *police* is affected by desyllabification (cf. Semiloff, 1973): /preɻd/, /pliɻs/. On the other hand, an emphatic pronunciation of *prayed* and *please* is affected by epenthesis:
>
> | | | |
> |---|---|---|
> | phonological form | //preɻd// | //pliɻz// |
> | epenthesis | /pəreɻd/ | /pəliɻz/ |

The point that phonetic plans pertain to careful pronunciations of the items involved can be argued in several ways, some of which may be notified here (see Linell, 1979).

(a) (As far as lexicalized items are concerned) the most careful pronunciation must (usually) be remembered since (often) it cannot be predicted from general phonological knowledge (rules) and knowledge of some *other* pronunciation of the item in question. "Reduced" (hypoarticulated) and "sharpened" (hyperarticulated) pronunciations, on the other hand, can often be derived from knowledge of the most careful pronunciation and general rules.

(b) Speakers seem to assign a special status to careful pronunciations; when they are asked to tell someone the pronunciation of a certain word, they tend to give it in its most articulated form (since giving a reduced rendition would leave out some knowledge they have).

(c) The primacy of careful pronunciations is also reflected in the perceptual reinterpretation of structure into defective input signals and in recall (see Section 19).

(d) People's intuitions about what is a phonologically permissible form in their language are clearly based upon careful pronunciations (Linell, 1979, Section 6.2.2). Similarly, speakers tend to assign phonemic status to a phonetic contrast only if it occurs at the level of careful pronunciation (ibid. Section 5.2.1).

(e) Note also that in practice, traditional, structuralist and generative discussions of the phonology of languages have (often implicitly) been restricted to careful pronunciations, which indicates that linguists have taken these to be *the* phonetic data of the various languages. Similarly, in phonetic models aimed at the explanation of "reduced" pronun-

ciations, especially vowel reduction, "target values" seem to be defined by careful pronunciations (cf. Lindblom, 1963).[9a]

(f) Finally, in applied contexts, e.g. language teaching, it seems practical to start out from careful pronunciations. Brown (1977), though mainly interested in making people more aware of what actual "informal" spoken English is like, still finds that in teaching spoken English one has to start out from " 'ideal', slow colloquial" as the ideal model.

To summarize, phonetic plans are relatively little-detailed phonetic structure specifications for careful pronunciations.

7. Phonetic plans and the problem of lexical representations

The reader may have noted that I did not list "account for lexical representations of expressions" among the requirements on phonological forms (Section 2). Such a requirement simply cannot be placed upon phonological forms in the sense of phonetic plans discussed here, and this for two simple reasons. On the one hand, while all linguistic expressions permitted within a language L naturally have their own phonetic plans, far from all expressions are lexicalized as ready-made items; instead, they must be constructed in the various communication situations. On the other hand, in order to construct (derive) the phonetic plans of the various expressions permitted by the language system we need not only phonological-phonetic information about those items which *are* lexicalized but also morphological, morphophonological and other information. Thus, determining the nature of lexical representations and finding out what phonetic plans are like are two quite distinct problems.

Since lexical items are not my concern in this paper, I will not speculate on their properties. However, reasonable considerations lead us to assume that lexical items are in fact given in terms of phonetic plans (i.e. plans of those items that *are* lexicalized) plus other grammatical and diacritic information (see Linell, 1979, especially chapter 4).

One of the basic mistakes of orthodox generative phonology (Chomsky & Halle, 1968) is the confusion of "phonological representation" with "lexical representation".[10] Their arguments for the superfluousness of phonemic representations may be interpreted as a denial that phonological forms (= phonetic plans) are necessary. But this position is indefensible. We have already seen conclusive arguments that the concept of phonological form (or representation) in the sense of plan for phonetic behavior is necessary in phonetic theory (Section 4). We have also argued that many types of evidence speak for the significance of surface phonemic contrasts (Section 5).

Interestingly enough, the orthodox generative-phonological discussion itself supports these view-points. Chomsky & Halle (1968), and other generativists, argue that one can derive "*the phonetic representation*" of a certain expression by applying phonological rules

[9a] The relationships between careful and reduced pronunciations are often discussed in terms of speech tempi (cf. the frequent use of terms like *lento*, *allegro*, etc.). However, the phonostylistic level intended and/or achieved in a particular utterance is primarily a function of the speaker's (conscious or unconscious) strive for carefulness and exactness in articulation, rather than a direct function of the rate of speech. Ultimately, carefulness in speech is dependent on social-psychological factors such as the speaker's interpretation of the communication situation as formal, familiar, intimate, etc. Hasegawa (1979) argues that there are partly different sets of articulatory rules for reduction and elaboration operating in casual speech vs fast speech.

[10] Chomsky & Halle (1968) do make a distinction – they introduce certain "readjustment rules" between the two levels of representation – but this difference must be regarded as minor in this context.

to the "underlying representation". However, it is nonsense to talk about *the* phonetic representation of a certain expression, for there are many, in fact infinitely many, phonetic representations of the same expression, since this expression may be pronounced in many ways and each pronunciation may be analyzed at many levels of accuracy (we would get "phonetic transcriptions" differing in narrowness). Obviously, these possible phonetic representations are not all equally significant, something that all phonologists, including generativists, intuitively know. It is therefore interesting to note that the particular phonetic representation that generativists discuss (derive) is in fact a little-detailed (roughly phonemic) representation of a careful pronunciation of the expression involved! Thus, the actual practice of generativists (who officially reject phonemic forms) provides indirect evidence for phonetic plans of the type discussed here.[11]

8. Plan construction and other kinds of planning: phonetic plans vs fully specified articulatory plans

The concepts of construction and execution need some further clarification.[12] In our conception, construction involves the choice of lexemes (with their (lexicalized) phonetic plans), the application of grammatical construction methods (including syntactic linearization and morphological operations) and the assignment of prosodic patterns, all of which is subordinated to communicative intentions. The output of plan construction is an expression plan. Thus, this aspect of utterance production is concerned with the linguistic identity of the utterances involved. A single change in a given expression plan would lead to a (potential) change of lexeme and/or grammatical construction ('potential' since not all phonotactically correct structures are actually utilized by the language). Therefore, it is supposed that the construction phase may involve active, sometimes partly consciously monitored choices. However, this apparently holds only for some aspects of any given utterance; many aspects, both grammatical operations and phonotactic adjustments, are automatized, some perhaps even partly probabilistic. There is no fully conscious, intended meaning before the message has been verbalized, i.e. before the utterance has been constructed (Linell, 1978). Furthermore, partial execution often occurs *before* the whole utterance, e.g. a sentence, has been constructed, although some advance planning is obligatory. In Section 15 I will return to the real-time relations between construction and execution.

[11] Chomsky & Halle (1968) sometimes discuss the putative functions of "*the* phonetic representation" in terms that may seem reminiscent of what I propose for phonetic plans and, even more so, for fully specified articulatory plans. Thus, "phonetic representations" allegedly represent independent elements in the perceptual reality of the speaker-hearer (pp. 25, 65, 298, 381), they may be viewed as (mental) instructions indicating how the physical articulatory system is to perform (p. 65), they represent "aspects of vocal tract behavior that are under the voluntary control of the speaker" (p. 381), etc.

[12] In recent years, speech production has often been discussed in terms of planning vs execution. However, there is no agreement as to how these concepts should be understood. Most scholars seem to work with one single, binary distinction (usually termed planning vs execution), although at least three concepts seem to be needed (i.e. construction, planning and execution). What different people mean by "planning" varies a great deal. Keller (1979) uses a narrow definition (more restrictive than my "construction"), including the selection of lexemes, grammatical constructions and intonational contours but excluding "morphophonemic, phonological and phonemic events" (p. 40). Hammarberg (1976), on the other hand, subsumes all aspects of language-specific programming (down to extrinsic allophones) under "planning", and thus reduces "execution" to cover only the purely "mechanical", universal aspects of the vocalization process.

One should distinguish a narrower concept of plan *construction* (just described) from a more comprehensive concept of planning (or programming) (in speech production as well as in other types of complex behavior). The concept of planning will be used to explicate all those aspects of speech behavior that are conventional or habitual to some degree, i.e. they are not fully predictable from the inherent, mechanical or other, constraints on the speech organs. It is obvious that the expression plan of an utterance, comprising the phonetic plans of the constituent word forms and a prosodic plan, specifies only some phonetic properties of the actual utterance. The actual execution of the specific utterance token can be explained only if we can pinpoint the *fully specified articulatory plan*, which must be specified down to all extrinsic-allophonic details. The various coarticulatory patterns of utterances do not arise simply as mechanical consequences of inertia or the like. They are partly natural and partly language-specific; though normally automatic and and unattended, they too belong to the language-specific programming necessary for producing idiomatic speech.[13]

This means that both construction and execution involve language-specific programming or planning. Note, however, that such programming does not necessarily presuppose active attention and voluntary control on the part of the speaker. On the contrary, learning to talk consists in no small measure in automatizing the production of many language-specific phonetic (and grammatical) aspects. However, we do assume that the construction phase involves much more of active, volitional planning than does the execution phase. Thus, although many aspects of expression plans follow from lexicalized properties and thus may be assumed to be programmed in advance, some active planning is necessary at the level of expression plans (what features of words and prosodics should be superimposed on the stream of vocalizations), and also in selecting the general level of articulatory accuracy (speaking style, phonostylistic level) to be used in the various parts of the utterance. On the other hand, the speaker need not construct specific articulatory plans for each utterance token. Rather, these plans automatically follow from the general habits of allophonic realization (hyper- and hypoarticulation). Since the biological constraints on the speech organs in themselves allow an almost infinite number of possible minute variations in vocal behavior, it is probably necessary that a great many details are regulated by automatized habits, thus allowing the speaker's volitional choices (active planning) to concentrate on just a few aspects.[14]

Thus, if we want to explain the production (and perception, cf. Section 17) of speech, we must postulate plans at different levels of specificity (phonetic plans vs fully specified articulatory plans). Note, by the way, that when we say that a certain invariant phonetic plan underlies all of the speaker's various manifestations of a certain word, this must be interpreted in a somewhat "indirect" way. Suppose a speaker intends to pronounce and does pronounce a certain word in a normal, somewhat "reduced" manner. Then I would not, of course, claim that he actually intends (consciously or subconsciously) to produce a maximally careful pronunciation (to which the phonetic plan refers). But he tacitly "knows" that the word he intends to produce has a certain phonological structure (= phonetic plan) and that this structure is constituted by some linguistically significant properties of the most articulated manifestation of the word.

Thus, the fully specified articulatory plan of a particular manifestation of a linguistic expression can be derived (i.e. predicted, "generated", calculated) given knowledge of the

[13] See Morin (1974), Hammarberg (1976), Tatham (1969) (the distinction between intrinsic and extrinsic allophones).

[14] Cf. Fowler *et al.* (1980) on "the degrees of freedom problem".

underlying expression plan (what words and prosodic patterns are intended) and of the general phonostylistic level aimed at. Technically, our phonological model would derive the fully specified articulatory plans by applying redundancy rules and rules for articulatory elaboration (hyperarticulation) and reduction (hypoarticulation).[15] Thus, we may suggest the following relations between the various plans underlying a possible (American) pronunciation of the word *winter*:

(2)[16]

phonological form (phonetic plan)	/wɪntər/	derivation by (perceptual) redundancy rules (for the term, see Linell 1979, p. 170)
fully specified articulatory plan of careful pronunciation	[wĩn · t^{h}ɚɹ]	(in reality much more detailed)
		derivation by articulatory reduction rules (e.g. nasal absorption, flapping, /r/ absorption)
fully specified articulatory plan of a particular "reduced" pronunciation	[wĩɾɚ]	(in relaity much more detailed)

[15] Thus, in terms of the phonological rule typology proposed in Linell (1979, chapter 10), perceptual redundancy rules, articulatory elaboration and reduction rules would belong to the execution phase, while morphological operations, including morphophonological rules proper, are clearly part of plan construction. Phonotactic rules occupy an intermediate position, although I am inclined to regard them as construction rules, since they may be seen as statements of conditions on the (coarse-grained) phonetic structure of phonetic plans (= the output of plan construction). On the other hand, many phonotactic rules are valid for the products of articulatory elaboration and reduction too. Some of these rules, notably lenitions, exist in a more constrained version at the level of phonetic plans, and are generalized in other speaking styles, where they function, as it were, as general articulatory reduction rulse (e.g. Swedish supradentalization, which operates only morph-internally in lento speech ([bu:ɖ] *bord* 'table, [bu:r dɯ:] *bor du* 'do you live') but is generalized to apply across word boundaries in allegro speech ([bu:ɖɯ·]), cf. Linell, 1979). Moreover, the distinction between phonotactic rules and perceptual redundancy rules is somewhat dubious.

Thus, while we can analyze aspects of production activities in terms of a conceptually clear distinction between construction and execution, we cannot expect all phonological rules to be monofunctional, i.e. belonging unequivocally to either construction or execution, unless, of course, well-known phonological regularities are split up onto several heterofunctional subrules (e.g. Swedish supradentalization as both a restricted phonotactic rule and a more general articulatory reduction rule). In other words, our typology of phonological rules is based on an analysis of phonological competence in terms of lexical storage, plan construction and execution (cf. lexicon, morphology/ including morphophonology/and phonology proper) (this theory is, in my opinion, akin to Dressler's (1977) so-called polycentristic theory), but some rules fulfil several functions simultaneously, or are open to competing analysis. This is to be expected, since languages are not fully integrated, closed systems, but instead they are partly open and subject to change (cf. Linell, 1979, p. 75, and p. 210).

[16] The conventional phonetic transcriptions are actually inadequate as a notation for articulatory plans (cf. Sections 8 and 11). For discussion of phonological rule categories, see Linell, 1979, chapter 10.

In Part II of this paper I shall be more concerned with the ontological and conceptual status of plan construction, planning vs execution, and of phonological forms and rules. However, it should be pointed out right now that rules must not be equated with performance processes. Instead, they should be seen as conditions on forms (plans) and operations or on the relations between forms. Furthermore, plans in phonology are not mental objects (mental sounds or the like). Rather, we must choose a more general and abstract characterization. A phonetic plan defines an equivalence class of behaviors, i.e. the class of (correct) pronunciations of the same linguistic unit. It serves this function by stating certain properties that must be preserved across instantiations. (Actually, some of the properties may vary with the particular phonostylistic level, but the variation is highly constrained by the rules of hyper- and hypoarticulation). Similarly, a fully specified articulatory plan delimits another equivalence class of behaviors, i.e. the more restricted class of pronunciations of the same linguistic item at a particular phonostylistic level. The various phonological rules (e.g. in (2)) state certain regular relations between coarse-grained (i.e. phonetic) plans and finer-grained (more specified) plans, but the relations are not of a causal type. It would be wrong to say, for example, that the phonetic plan /wɪntər/ is the cause of the fully specified plan [wɪ̈ɾ̈ɚ].

9. A note on the contents of phonetic plans

At this point I would like to insert a few comments on the linguistic contents of plans, in particular phonetic plans. Following a long tradition I use a notation based upon letters which stand as symbols for phonetic segments. Apart from the obvious descriptive convenience of using alphabetic letters, does this amount to a theoretical claim that plans are actually organized in terms of segments, i.e. linearly ordered, discrete and invariant bundles of simultaneous features? Most phonologists would clearly maintain that this is indeed the case, in spite of the fact that phonetic behavior is varying and continuous rather than composed of discrete elements. The segment is indispensable to phonology in the opinion of many scholars. Hammarberg (1976) even argues that the very concept of coarticulation, i.e. that phonetics is *not* segment-like, shows that there are abstract segments, because coarticulation must be seen as "the influence of one speech segment upon another". This reasoning[17] looks like something of a logical somersault, since the phenomenon of coarticulation might *prima facie* invite a conception, according to which several linguistic components are simultaneously present; their overlapping would be basic and natural rather than considered to represent a *destruction* of a purportedly existing underlying reality of abstract and serially ordered segments. It would therefore be motivated to inquire into the logic of the phonological-phonetic segment theory.

The basic observation concerns the fact that the sound gestalts of the words of a particular language are normally distinct from each other in systematic ways. Thus, e.g. the sound gestalt of *bit* is different from those of *pit, bid, bet, beat*, etc. Moreover, the types of differences seem to recur in other words. It is therefore reasonable to assume that the sound gestalts are different by virtue of being composed of, or at least analyzable in, certain distinct, categorical elements that can combine in different patterns. That there are categorical invariants underlying speech may be argued on different grounds. For one thing, perception in general is categorical. If we do not want to assume absolutely invariant, stable elements, we might invoke the concept of perceptual prototypes (focal instances on con-

[17] This reasoning is of course not confined to Hammarberg but is implicit in the generally presupposed frame of reference in contemporary phonetic theory. For similar statements, see Kent & Minifie (1977, p. 116).

tinuous dimensions, cf. Rosch, 1975). Furthermore, there seems to be evidence for relatively invariant speech gestures at the production side too (cf. below).

The next step in the traditional argumentation assumes that the basic elements are serially ordered "thing-like" *segments* drawn from a limited repertoire (of phonemes), i.e. segments defined by the conditions of linearity, discreteness (non-overlapping) and invariance. It is almost certain that this assumption has been reinforced by the existence of the alphabetic writing systems used in the cultures to which linguists have belonged. In general, present-day linguistic theory is heavily influenced by the linguists' traditional preoccupation with written language (Olson, 1977; Linell, in preparation), and the case of phonological theory is an obvious example. Indeed, phonologists usually treat phonological forms as if they were just strings of written symbols, and phonological rules as if they involved replacements of symbols (Bailey, 1979).[18]

Now, there is indeed a good deal of evidence that serially ordered elements are basic structures in phonetic behavior (see summary in Linell, 1979, Section 3.7). Accordingly, Fowler *et al.* (1980), who argue against segments as static parts underlying speech, still endorse the view that "linguistic units are *qualitatively separate and serially ordered* (but not discrete), *dynamic* and *context-free*" (op. cit. p. 375). In other words, it seems obvious that vowels and consonants are real and significant units in speech. However, it is not necessary to assume that vowels and consonants are linearly ordered, non-overlapping segments (in a narrow sense). Another important caveat is that it is hardly plausible that vowels and consonants ("segments" in a less strict sense) are the *only* significant units of dynamic complex gestures in speech. In fact, syllables and whole word gestalts are also plausible candidates (perhaps in particular as regards early child language). Moreover, we might be forced to use not only segments and segment-based features but also some suprasegmental components, perhaps as in Firthian phonology or the more recent theory of autosegmental phonology (Goldsmith, 1976). In yet other respects we do not know how to characterize the basic gestural components in speech. Obviously, we are dealing with movements towards certain acoustic target effects (Section 14), but are the movements as such the primary units (and the targets only derived values that are asymptotically approached), or are the targets primary and the movements predictable means for approaching the targets? Furthermore, we do not know how to represent the time dimension in phonetic and/or articulatory plans (cf. Section 11).

10. On the psychological and social status of plans

Anyone who is familiar with the history of phonology knows that notions like plans or intentions have been discussed in connection with phonological forms several times before. Some scholars have strongly objected to proposals similar to mine. Thus, Trubetzkoy (1958, p. 37) rejected the notion of 'Lautabsicht' ("intended sounds") as a possible interpretation

[18] In addition, one might single out a third step in the (sometimes only implicit) argumentation for abstract segments in phonological representations, i.e. that these phonological segments must be elements of morpheme-invariant representations. This argument, which is based on the erroneous assumption that morphemes are the only primary, lexical expression units, will not be probed in this paper (see Linell, 1979, especially chapter 12). However, it may be seen as something of a logical continuation of the above-mentioned *second* step; that is, since phonological forms are necessarily segment-based and segments must be abstract (i.e. they are not part of overt behavior), we might just as well make them so abstract (non-phonetic) that they can serve as elements of morpheme-invariant representations.

of the notion of phonological form; he argued that the intended sounds are not phonemes but allophones. In our wording, he would probably think of Lautabsichte as fully specified articulatory plans.

However, terms like "plan" or "intention" may perfectly well be used when we consider only the linguistically relevant properties of the intended behavior. In fact, this is the most reasonable and straightforward interpretation; the linguistic identity of the expressions used (and the general speaking style) is what matters in communication and hence also in speakers' active planning of communicative acts. Maybe a comparison with another example of human intentional behavior would be appropriate here. Suppose a tennis player intends to pass his opponent by striking a backhand cross. Then we would conceive of his plan for this act as including the specifications that could distinguish this kind of shot from other possible shots that the player might choose instead. It would be odd, however, to conceive of his plan as specifying all the innervations needed for bringing about all the required movements of muscles, joints etc given that the player has a certain point of departure for the movements that realise his action.

Thus, I would maintain that the interpretation of phonological forms as phonetic plans is both natural and necessary, provided that we want to explain the role of phonological structure in actual speech performance, which is what all phonologists who are genuinely interested in explanation rather than mere description should ultimately aim at. The alternative interpretations of Trubetzkoy and other classical structuralists would not do. It is common among structuralists to contend that phonemes are linguistic, not psychological, concepts. It is argued that their identity should be sought solely in their linguistic "function". Then, if we ask for their ontological status, we are told to think of them as social facts rather than psychological entities:

> Das Phonem ist vor allem ein funktioneller Begriff, der hinsichtlich seiner Funktion definiert werden muss." (Trubetzkoy, 1958, p. 38)
>
> Da das Phonem zum Sprachgebilde gehört und das Sprachgebilde eine soziale Institution ist, ist das Phonem eben ein Wert und besitzt dieselbe Art von Existenz wie alle Werte" (ibid. p. 41)

Though Trubetzkoy here discusses "phonemes", one may assume that his arguments would also be valid for the phonological forms of words and utterances. It is of course perfectly possible to regard phonological forms as "functional information" or "social values" *tout court*. But such characterizations are not sufficient by themselves. If phonemes and phonological forms are "functional", how, by what means, do they fulfil their functions in speech performance, and how do they acquire their function? If they are irreducible social facts, how do they interact in real people's linguistic performance? Such questions must be at least provisionally answered, if linguistic theory is to achieve both psychological and social validity. Therefore, we would think of linguistic items as socially shared dispositions to behave, and expression plans as intentions to produce phonetic acts defined by certain socially observable, linguistic conditions (rules). Language considered as part of a person's communicative competence is a psychological entity, but it can grow and get established only through social interaction.[19]

[19] Note that even Saussure modified the Durkheimian conception of linguistic items as irreducible social facts; though *langue* is seen as a social phenomenon, linguistic signs are characterized as psychological entities.

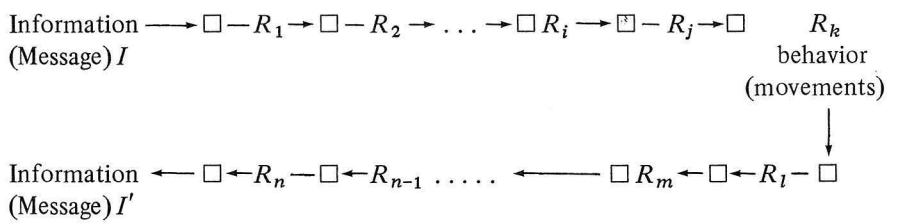
Part II

In the first part of this paper I tried to develop an interpretation of the notion of "the phonological form of the expression X" that relates rather directly to what speakers actually do in their phonetic behavior. In Part II, I will also pay attention to listeners' activities. However, I will also further analyze the nature of plans, in particular plans for linguistic behavior. I will do this by comparing the dominant paradigm of information processing with their *translation* theories of speech production and perception to another kind of attempt at understanding mental activity, according to which the agent "*directly*" interacts with the outer observable reality.[20]

11. Translation theories of speech behavior and activities in the real world

Contemporary text-books on cognitive psychology and information processing usually offer *translation* models of the mental activities involved in e.g. linguistic performance. In communication, a message is assumed to be translated by a complex internal machinery from one form (representation) into another as in the abstract diagram of (3):

(3)

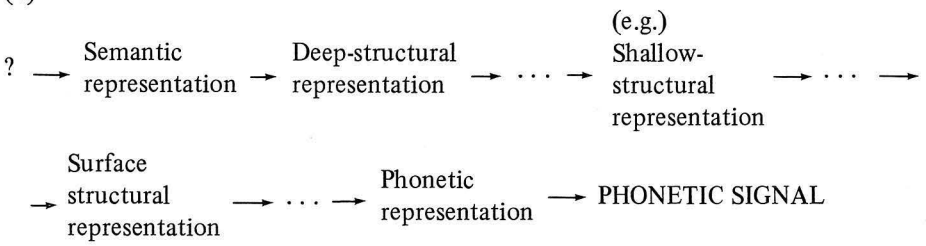


In each box (□) some kind of "processing" ("translations", "transformations") is carried out. $R_1, R_2 \ldots . R_n$ are a series of "representations" of the information processed. $I \approx I'$ if no harmful noise intervenes during the translation process.

If we take this model at face value, there is very little room for constructive or exploratory activity on part of the behaver (speaker-listener). Rather, a message is automatically processed through a series of abstract representations, which, according to those psycholinguists who believe in current linguistic competence theories, may be equated with various purportedly "linguistically significant" representations. Thus, the production processes may be pictured as a linear sequence of translations of the same utterance via representations such as the following:

[20] The term *translation theory* has been used by several authors, e.g. Garrett (1975). The metatheoretical critique of it seems to have been developed independently by several people, including Neisser (1976) (perception in general), Linell (1978) (utterance production) and Fowler *et al.* (1980) (speech production). See also Öhman (1975*b*), Norman & Bobrow (1976) and Olsen (1979). The term *direct* (non-translational) theory of speech production has been used by Fowler *et al.* (op cit, p. 382). Since "translation" in this context implies something like (iterative) mechanical transfers of the same content across different modes of representation (coding), it has rather little in common with actual translation of literature between languages (on translation, see Steiner, 1975).

(4)

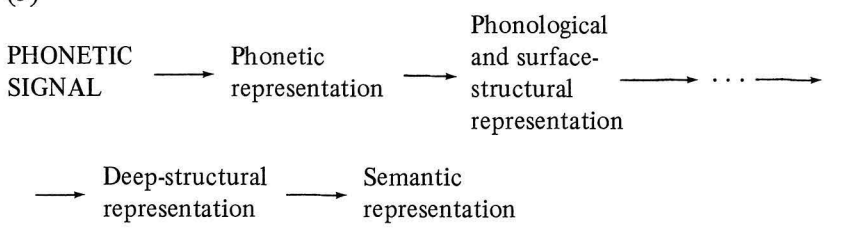


Exactly what "stages" are included naturally depends on the particular theory espoused. In general, it is clear that early ("remote") representations may be very different from late ("superficial") representations though they still "represent" the same message unit (usually a sentence). I have shown elsewhere (Linell, 1978) that such a theory is untenable.

Instead, we propose a theory based upon the distinction between construction and execution. Since most utterances are not stored as ready-made items in the memory, they must be constructed. But what the speaker constructs is not some more or less mystical, "remote" representation but the actual utterance itself. Thus, plan construction pertains "directly" to the observable words and prosodic patterns which are also actually spoken. (We assume that the utterance must not only be constructed but also properly executed.) An utterance plan is a plan for dynamic behavior, not a sequence of abstract "things" (states). To speak is to behave phonetically in the real world, and to construct an utterance is to plan actively a sequence of observable phonetic events. We must admit, of course, that before the whole utterance has been constructed, there must be preceding stages of pre-planning, parallel and sequential planning of constituents and single aspects of utterances, but the available evidence shows that a translation theory cannot adequately capture this process (Linell, 1978).

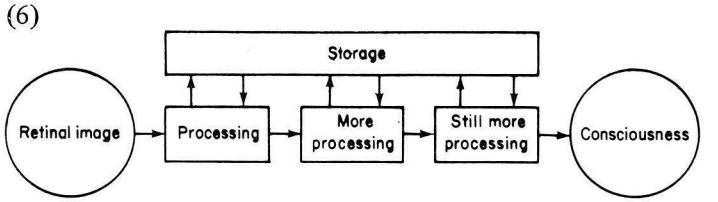
Similarly, there are translation theories for perception and understanding. In the most naive version, this is simply the reversal of (4):

(5)



The addition of some analysis-by-synthesis mechanisms does not change this general outlook. Theories of this type belong to the class of linear sequential "information processing' models of perception criticized by Neisser (1976, pp. 16–7, 113).

(6)



(From Neisser, p. 17).

Neisser criticizes, e.g. the theory of vision, according to which we see (or, as it is nowadays put, "process") retinal images. The same kind of arguments can be raised against similar theories of other sense modalities. (The reader is referred to Neisser's book, which contains much of interest for linguists). To put it very bluntly, in speech perception we do not hear mental percepts of phonetic signals ("phonetic representations") or phonological, syntactic or semantic representations of utterances, but we hear the actual utterances themselves and understand their meanings and, perhaps, their linguistic properties.

To think in terms of translation theories of mental processing means that one easily gets caught in the trap of dualism; one is forced to postulate a ghost in the machine (Ryle, 1949), a homunculus who perceives retinal images or phonetic representations. Ryle showed the paramechanism involved in this kind of modelling mental activities. This (para)mechanism is characteristic of a long, mainly American tradition in psychology, comprising behaviorist multi-stage stimulus–response psychology, cybernetics, the Katz (1964) conception of mentalist models of language, and information processing models of contemporary cognitive psychology. Despite some obvious differences, all these seem to build upon the same, largely unquestioned ways of modelling man (the speaker, listener, behaver) not as someone who does things but as something to which things happen, not as someone who acts meaningfully in social contexts but as something like an information-processing system functioning in a social vacuum.

12. Time in models of speech performance

Linguistic behavior is distributed in time. We can think of linguistic structure in terms of modulations on a stream of vocal behavior.[21] Since time is always a relevant dimension, it is interesting to examine how it is treated in linguistic models, in particular in models of speech production and perception.

Most models of utterance production and understanding make use of various linguistic categories and units, rules, levels of representation etc. It is therefore important to call attention to the fact that a grammar as such, e.g. a generative transformational grammar, is *not* a model of time-bound behavioral processes but an abstract timeless model of relations between different levels of representation. Chomsky has repeatedly emphasized that it is a serious category mistake not to observe "the difference in logical character between performance and competence" (1972, p. 71). Nevertheless, psycholinguists have been more or less forced to impose some kind of behavioral interpretation upon grammars, when they have tried to explain performance by somehow "incorporating a competence model" (cf. Chomsky & Halle, 1968, p. 110). Thus, they do in fact work with the well-known levels of linguistic representation within typical translation models (Section 11).

When one considers the role of time within translation models of speech production and perception, one should distinguish the treatment of temporal relations *between representations* from those *within representations*.

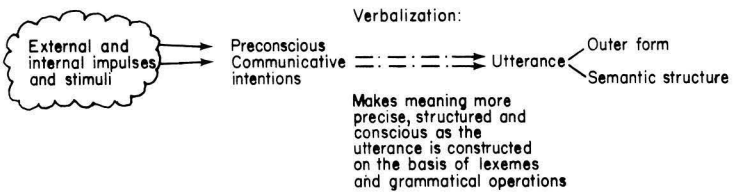
It should be clear (cf. Section 11) that a translation model crucially involves the assumption of temporal relations between the various representations (e.g. in (4)); a semantic representation or intended meaning is successively translated into various forms of representation up to the production of overt phonetic behavior. This involves a category mistake; means–end relations are erroneously treated as mental states arising in a temporal

[21] We disregard here the accompanying non-vocal (non-verbal) behavior.

sequence. The semantics, grammar and phonology of a given utterance are the speaker's means (at various levels) for attaining certain communicative goals. Even if there are significant dependencies between, say, the semantics and the outer form of a given sentence (and the grammar-semantics relation is more or less a means–end relation), the meaning does not exist *before* the outer form, nor is it the cause of the latter. Rather, semantics and grammar are two aspects that both result from the construction of the sentence in question. Similarly, if one uses an eye wink for flirting with someone else, this does not mean that the meaning of flirtation (or, say, 'I like you, why don't you join me') is present *first* and is only *then* translated into a wink. Rather, flirtation as a content evolves simultaneously with the movement "expressing" this content or intent.

The obvious objection to this is that any act must involve some intentions, some planning or "information processing", before the actual execution in overt behavior. This is undoubtedly true, but the "planning" prior to execution is not of the kind that most linguistically based translation models assume. There is no complete intended meaning before verbalization, prior to the overt realization in behavior (unless behavior is tacitly rehearsed, Section 15). Thus, as regards language, I have argued elsewhere (Linell, 1978) that the production of an utterance must be seen as an activity in which communicative intentions, which are partly imprecise, vague and preconscious at the outset, become gradually more structured, enriched, precise and conscious through the verbalization process itself:

(7)



"An act is continually becoming; its coherence emerges over time." (Fowler, 1980, p. 110).[22]

Let us now advert to the other question concerning the treatment of time within linguistic representations, e.g. within the syntactic representation of a given sentence or the "phonological form" of a certain word. It then turns out that there is *no* timing involved except for the pure linear order of elements (words or morphemes in grammar, syllables or segments in phonology), and this serial order is spatial rather than temporal. By focusing on this property of linguistic (and especially phonetic) models of performance, Fowler (1980) has dubbed these models of *extrinsic timing*. The class of extrinsic timing models seems to be extensionally identical to the class of translation models.

According to the traditional conception, a phonological representation is organized as a linear sequence of discrete segments. "Segments in a planned sequence are discrete in the sense that (abstractly stated), their boundaries are straight lines perpendicular to the time axis, so that the terminus of one segment is the beginning of the next segment" (Fowler, 1980, p. 116). Each segment is a bundle of simultaneous features. Such a segment sequence is thus a basically spatial (rather than temporal) organization of thing-like phonological units arranged in a before–after sequence analogous to the left–right sequence of conventional orthography (and conventional phonetic notation). The phonetic processes of coarticulation

[22] Similar arguments can be found in many places, cf. Sperling (1977) (on Merleau-Ponty), Danks (1977).

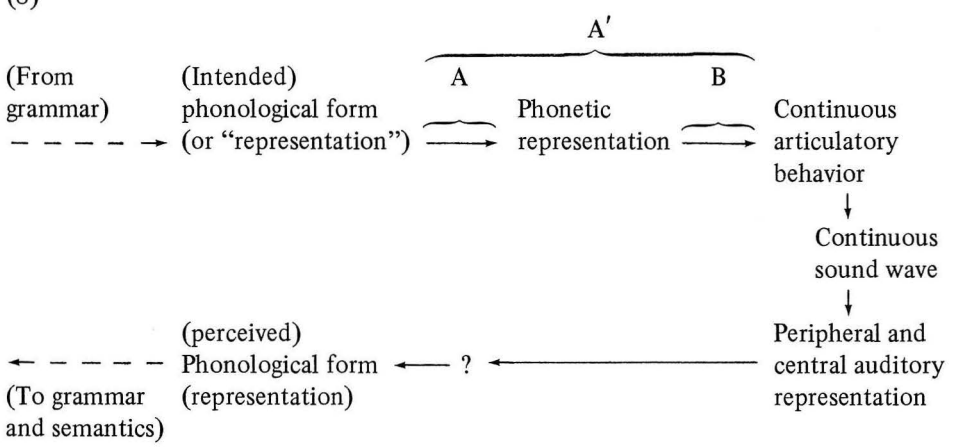
are thus viewed as a translation of a *spatial* configuration into a time-bound varying stream of behavior. The view of plans as spatial configurations is associated with theories of memory, according to which memory traces are static, simultaneously existing but spatially differentiated structures. However, "memory traces" might alternatively be seen as simultaneously existing but temporally organized patterns of ongoing neural activity.

Fowler (1977, 1980) argues, convincingly in my opinion, that timing must somehow be represented in phonetic and/or articulatory plans,[23] i.e. it should be *intrinsic* to those representations of or plans for linguistic behavior that are activated or constructed and used in speech production and perception. A simple, spatially organized order of segments is inadequate for several reasons, some of which have already been hinted at (Section 9) or will be mentioned below (Section 13). In addition, one may notice the following problems: (1) the timing adaptations needed for different stress patterns and tempi must be somehow represented in plans, and (2) at least some segment types have a clear, internal temporal organization, i.e. they are not simple bundles of simultaneous features (e.g. prenasalized stops) (Fowler, 1980, p. 117).

13. Phonology in translation theories of speech production and perception

I will now narrow down my discussion of translation theories, as opposed to "direct" theories, by concentrating on the phonological-phonetic level, or stage, of the translation processes of (4) and (5). This we can diagram as follows:

(8)

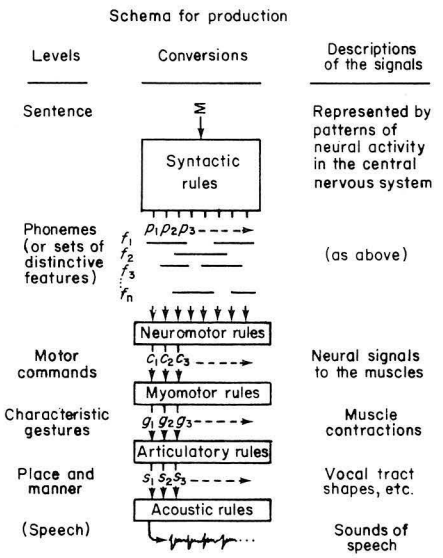


Thus, the phonological stage starts with the speaker's intended phonological representation ("underlying phonological representation"), which is transformed in various steps into continuous articulatory behavior, which in its turn results in acoustic patterns. In the listener, these patterns are received and analyzed in ways that are usually admitted to be poorly understood. However, the phonological-phonetic stage may be said to terminate in a perceived phonological form, which is hopefully identical to the speaker's intended form. As Liberman *et al.* (1967, p. 453) put it, "the listener uses the inconstant sound as a basis for finding his way back to the articulatory gestures that produced it and thence, as it were, to the speaker's intent".

Naturally, there are different opinions among theoreticians as to how exactly the various

[23] See also Coates (1980).

hypothetical constructs should be understood. For example, there are vast disagreements concerning the level of abstractness of phonological forms. However, the fundamentals are more or less the same. One usually thinks of phonological forms as strings of abstract segments (or feature complexes), and these "abstract phonological segments are premotor cognitive entities that ⟨ . . . ⟩ can serve both as outputs of perceptual processing and as inputs to an articulatory mechanism" (Fowler *et al.*, 1980, p. 376). Some would assume the existence of a phonetic representation as an intermediate stage in the production process. If so, processes labelled A in (8) represent the counterpart of the phonological derivation of a generative phonology, whereas (B) comprises the "phonetic" level of coarticulation. Other theoreticians might prefer to speak of (A′) as one block of coarticulatory and other phonetic processes, or as a series of translations (or "conversions") performed by different kinds of "rules" such as "neuromotor rules", "myomotor rules", and "articulatory rules" (Liberman *et al.*, 1967, see Fig. 9).



(9) Schematic representations of assumed stages in speech production acc to Liberman *et al.* (1967, p. 445).

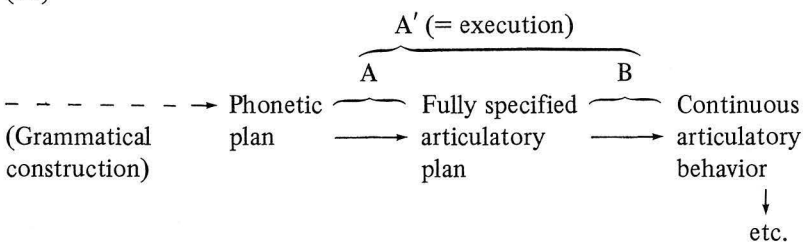
Translation theories of speech production and perception represent the prevailing method of reconciling two different, purportedly real modes of existence of speech sounds, i.e. the phonological strings of discrete, static and context-free segments and the phonetic, continuous and dynamic behavior. As Fowler *et al.* (op. cit.) put it, this means that speech production implies a destruction of the clear-cut phonological structure, while speech perception involves a reconstruction of it. The first-mentioned process, that of coarticulation, is described in this way by Daniloff & Hammarberg (1973), Hammarberg (1976), Kent & Minifie (1977) and many others. The phonetic manifestation process is characterized as "the assignment of phonetic effects to phonological causes" (Hammarberg, 1976, p. 356), or as a "conversion" or "translation" of phonological features "to articulatory transitions" (Kent & Minifie, 1977, p. 131). "We assume that at some neural level the units of speech encoding are the canonical forms. They are invariant, ideal, uncoarticulated target forms. Either the strategies of encoding or the actual process of encoding can introduce variability

into canonical forms such as occurs in coarticulation of speech segments" (Daniloff & Hammarberg, 1973, p. 240). Liberman *et al.* (1967, p. 447) argue that most of the restructuring, i.e. the destruction of phonological structure, occurs at the stage of "conversion from muscle contraction to vocal tract shapes by way of Articulatory Rules" (cf. (9)), but other possibilities are also open within translation theories. The predominant thinking may perhaps be summarized by Hammarberg's (op. cit.) contention that the notion of coarticulation logically presupposes the existence of abstract phonological segments (cf. Section 9). In short, the basic hypothesis is that phonemes are translated into extrinsic allophones (processes A), and these allophones are realized as continuous behavior (B).[24] Linguistic processes (A) are typically pictured as "replacements of things" (Bailey), although in our theory we should prefer, with Bailey, to see linguistic structure as modulations of behavior.

Translation theories of speech perception are legion in standard works on speech perception. Thus, Studdert-Kennedy (1976) describes the stage of "phonetic perception" as "the transformation of a more-or-less continuous acoustic signal into what may be described as a sequence of discrete phonetic symbols" (p. 243).

The critical reader will now surely ask: In what way is your theory of phonetic plans and construction vs execution different from a translation theory? It is clear that phonological forms (alias phonetic plans) are assumed to be much more "concrete" then in e.g. orthodox generative phonology, but barring that, what prevents us from interpreting the concepts introduced in Part I in terms of stages and translations, so that we get (10) simply by substituting some terms in (the upper part) of (8)?:

(10)



[24] However, Hammarberg's (1976) paper is not quite consistent and even seems partly self-contradictory. At the beginning the author argues distinctly in favor of the idea of phonological forms as sequences of abstract segments and the related idea of coarticulation as transforming this structure (thereby destroying it) into continuous phonetic behavior. Later (approximately from p. 359 onwards) he demonstrates his cognizance of the dualism inherent in this view and makes some attempts at circumventing it. In particular, he argues against a sharp distinction between planning and execution, and objects to the conventional view of coarticulation as pertaining solely to performance, i.e. execution. Coarticulation is always subject to numerous language-specific traits (cf. section 8), and the processes involved are therefore phonological rather than phonetic in Hammarberg's opinion. It is mainly on this point that he deviates from the prevailing theories. His general framework, however, is clearly a translation theory of speech production (i.e. an extrinsic timing model, cf. Fowler, 1980). He postulates a block of phonological, mental processes (including coarticulatory processes) preceding and even causally impinging on a block of phonetic, physiological processes. Such a view is awkward or even unacceptable; there is only one phenomenon, i.e. phonetic behavior, which is subject to various, language-specific conditions. Phonology is concerned with these language-specific constraints on the use of phonetic mechanisms. (In our model most of coarticulation would be handled by articulatory (reduction) rules, cf. section 8 and Linell, 1979, chapter 10).

It must be admitted that it is very easy to slip into this way of thinking. In fact, some of my earlier writings might suggest such an interpretation (e.g. Linell, 1979). But I believe that an action theory involving the construction and execution of plans need not be interpreted in this way. Maybe, it is fair to say that many people, including myself, are now aware of the inadequacies of translation theories for complex behavior, while we are still struggling to find a more plausible conception. The purpose of the rest of this paper is to further explain some features of a direct theory of speech production and perception.

14. The function of plans in speech production

In Part I I proposed that phonological forms could be assigned a real function in both speech production and speech perception, namely as plans for language-specific phonetic behavior and for language-specific exploratory activities in perception. I will postpone the discussion of the role of these plans in perception to Sections 16–20. Presently I will continue to focus on speech production. How should we conceive of a plan for phonetic behavior?

It is not necessary to espouse a translation theory of speech production, according to which the phonetic plan, conceived of as a chain of concatenated abstract, mental objects (mental sounds or, perhaps, neural states or neural commands), is translated into vocal behavior. It would also be misleading to say that the speaker first constructs a categorial plan and then executes it in behavior. Therefore, the term "plan construction" is something of a misnomer; what the speaker constructs is utterances. It would also be wrong to talk about plans and behavior in terms of causes and effects. What the speaker produces is not the plan and then the behavior; he simply produces vocal behavior, and he does so in a structured, planned or intended manner. It is typical of correct language-specific vocal behavior that it fits certain linguistic rules or conditions, i.e. those rules that are defined by the plans.

The relation between plans on different levels of generality should not be stated in causal terms. If the sound gestalt of a certain word in careful pronunciation can be characterized in certain phonological terms, i.e. as a phonetic plan, then we can predict according to more-or-less general rules how this word will be pronounced at various other phonostylistic levels (Section 8); the features of the fully specified articulatory plans are implied by or predictable from the phonetic plan. However, this only means that plans are more-or-less specified and that the specification of some features ("predictable" features) follow from the specification of certain other features ("inherent" features of phonological forms); it does not mean that some plans are caused by other plans. Similarly, behavior is not the "phonetic effects of phonological causes" (Hammarberg). Rather, behavior is intentionally produced or caused by the speaker himself.

Plans are specifications of structure, or conditions, imposed on behavior. In the case of speech, we can image a carrier wave of vocal behavior characterized by, let us say, "the special mode of breathing, the phonation mode of laryngeal adjustment, suprasegmental timing constraints, the near alternation of consonants and vowels, etc." (Fowler *et al.*, 1980, p. 386). Phonetic and articulatory plans specify features that are superimposed on this carrier. Thus, the traditional area of segmental phonology seems to correspond to at least two sorts of such superimposed modifications, i.e. those resulting from, on the one hand, the slow, continuous vowel gestures responsible for the global shaping of the vocal tract, and, on the other hand, the fast, quasi-discrete consonant gestures leading to various local obstructions of the tract. In Fowler's words, segments correspond to qualitatively separate and overlapping gestures.

To sum up, phonological forms constitute language-specific conditions or modulations

on the continuous stream of vocalizations rather than "mental objects" fabricated by a cognitive mechanism underlying speech. With Fowler *et al.* (op. cit.), who discuss speech production as a special case of intentional motor behavior, we can perhaps talk about (phonetic and articulatory) plans as "co-ordinative structures" governing complex muscular collectives, and "co-ordination is to be understood not in terms of the "issuing of commands", but rather in terms of the "arising or specification of constraints" " (p. 394).[25]

15. Some mistaken arguments in favor of a translation theory of speech production

There are some arguments that could be used in favor of a translation-theoretical interpretation of our plan construction-and-execution model (cf. (10)), and I will briefly consider three such arguments, which have to do with (a) anticipatory planning, (b) silent speech, and (c) neurophysiological reality.

The study of spontaneous speech provides several types of evidence that speakers are normally planning their speech several syllables or words ahead. That is, when they pronounce one part of the intended utterance, they have already planned many aspects of what will come further ahead in their utterance. Garrett (1975) argues that a rather detailed phonological planning may comprise some five-to-seven syllables ahead, whereas a more general planning pertaining to grammar and semantics reaches much further into the future of the utterance to be produced.

What is planned to appear later on sometimes interferes with what is pronounced earlier. This is the case in many slips of the tongue. It is particularly significant that typically these errors can be described as anticipations or permutations applying to underlying expression plans (or phonetic plans), and that the resulting rearranged plans only later seem to be subjected to phonotactic and coarticulatory readjustments. For example, Fromkin's (1973) corpus of English slips contains many items like (11) involving the English plural morpheme:

(11) *cow tracks* [træks] ⟶ *track cows* [kæwz] (op cit., p. 27)

It is conceivable that the speaker first generates an expression plan, then an interchange of *cow* and *track* causes the plural element to become adjacent to *cow* instead of the intended *track*, and then, finally, perhaps in connection with the execution, the morpheme is adjusted

[25] The action theory of speech production (as described by Fowler *et al.*, 1980) has evoked a considerable amount of interest, and, to some extent, skepticism among phoneticians. Witness, e.g. the debate at the 1979 Copenhagen international phonetics congress (cf. Lubker *et al.*, 1980). The skepticism that has emerged so far seems largely to be concerned with the possibilities of empirically falsifying an action theory of speech performance (e.g. op cit., p. 280). A couple of comments may be appropriate here.

First, we are here dealing with the choice of a very basic, theoretical or philosophical framework, the most general categories within which we want to think, and we cannot expect such a framework to be amenable to absolute empirical falsification. Such frameworks may (and must) of course be critically discussed and assessed for fruitfulness, but they cannot be conclusively falsified (let alone proved).

Secondly, the adoption of a teleological theory of behavior (like action theory) does not eliminate the necessity of searching for explanations of other kinds and at other levels, e.g. within causal, neurophysiological theories. The point made by proponents of action theories is simply that complex behavior like speech cannot be fully understood without a teleological perspective (Section 3). It is not denied that some aspects of the behavior call for other types of explanations.

to the new environment, i.e. the plural element is turned into a voiced segment (note that it would have been voiceless in *tracks* and that [kæws] would have been a phonotactically correct English word form (cf. *mouse*) although it is wrong as a plural of /kæw/). Or consider the following speech error (from Langacker, 1972, p. 247):

(12) *Freudian slip* ⟶ *Shreudian flip*

Here, it would seem that /f/ and /s/ of the underlying plans are interchanged, and then /s/ turns into /ʃ/ in the new environment, because /sr/ is an impossible initial cluster in English.

Slips involving permutations or anticipations of future units never occur across hesitation pauses. This fact supports a common, intuitive interpretation of hesitations; in fluent phases of speech production planning is progressing smoothly, well ahead of execution, but now and then the speaker fails to find the right words, hesitates, and produces (usually filled) pauses, and in such cases execution seems to catch up with plan construction.

Thus, speech errors (together with other types of evidence) confirm that speech is a form of complex motor behavior involving anticipatory planning as a characteristic feature. Furthermore, only things that are contained in phonological forms (phonetic plans), i.e. aspects which are presumably more actively planned and controlled, tend to be subject to permutations and other errors, whereas more automatic aspects, e.g. coarticution and phonotactic constraints, never seem to be violated in such errors. Furthermore, there seems to be evidence for a certain, at least partial, temporal precedence of phonetic plans (expression plans) with respect to fully specified articulatory plans and their execution. This much seems to follow from the available facts. However, this does not allow us to conclude that there are thing-like plans consisting of strings of mental segments, which can be scrambled in slips and then transformed into behavior. The only thing we have to assume is that there are several different processes going on simultaneously in the speaker's planning activities and that some of these processes are anticipatory, i.e. they are concerned with aspects of the future behavioral stream. Furthermore, it seems likely that the more coarse-grained planning spans a larger part, i.e. it goes further ahead, while the finer-grained plans are more restrictive in their anticipation. Note, however, that also at the level of detailed coarticulation there are anticipatory phenomena, e.g. so-called regressive assimilations. Thus, there is parallel processing at several levels of the planning activities, but the activities do not give rise to plans taken as abstract mental structures; instead, they end up in behavior itself, i.e. in a planned, linguistically structured behavioral stream.

The next argument is concerned with what happens when we "speak for ourselves", i.e. when we "think" explicitly in terms of linguistic constructions so that we can hear "in our inner ear" what we say to ourselves. It might be argued that this involves generating the plans for the corresponding utterances but not executing them. If so, silent speech would be construction (planning) without execution, and, therefore, the existence of silent speech would provide evidence that construction and execution are indeed distinct stages in performance.

However, this argument is unsound too. Although it seems that when we talk to ourselves, we do only parts of what we do when we speak aloud, and that such inner speech is similar to normal social speech in as far as construction is concerned, it would be counter-intuitive to argue that inner speech is only planning. In our terms, there is some kind of "inner execution" too. Certainly, what we hear "in our inner ear" is not abstract plans, mental sounds or the like; rather, inner speech "sounds like" real executed speech. When we imagine something being said, we do not imagine plans but potential behavior. I will return to the problem of phonetic imagery in Section 20.

We can also approach the problem of inner speech from the vantage-point of normal social speech. The theoretical position that we are currently criticizing would then seem to amount to the proposal that normal speech can be analyzed in terms of distinct processes of plan construction and execution, possibly distributed in time so that execution lags behind planning by some units of time, and that the first stage, that of plan construction, is identical to what occurs in silent speech. But this is an impossible theory; clearly, in normal speech we do not first produce silent speech to ourselves and only then pronounce the same thing aloud. Thus, our distinction of plan construction and execution is completely different from whatever distinguishes silent speech and overt talk; in fact, it must not be conceived of in terms of stages in a sequence of translation processes at all.

Finally, let us turn to the neurophysiological argument, which would go something like this. Before the various speech organs of the speaker can be moved in the ways desired, they must be commanded to do so by neurophysiological means. Thus, the articulatory behavior is related to neural commands as effect to cause. Now, we claim psychological reality for phonetic and articulatory plans, and it would seem natural to assume that there is a basis for the reality in the neurophysiological processes that precede overt speech. Otherwise, it might be argued, one can hardly claim that plans fulfil a real function in speech production (and perception).

In fact, this is the standard view underlying most theories of speech production and perception; invariant phonological forms coded in terms of patterns of neural commands are translated into coarticulated speech (see Section 13). However, it is precisely this theory that has been called into question by so many phoneticians in the last decade. Peripheral motor commands are, by and large, *not* invariant according to electromyographical findings. In fact, peripheral commands *must* be variable because of all the possible differences in the initial conditions of the speech organs, in the articulatory contexts and in non-speech conditions (e.g. food in the mouth, various obstructions for individual organs). Moreover, the movements of the articulators are not only the result of muscular forces but also of reactive forces (Fowler *et al.*, 1980, p. 392). On the other hand, the intended *results* must be relatively invariant. Speech is "output-oriented"; "the essence of the speech production process is not an inefficient response to invariant central signals, but an elegantly controlled variability of response to the demand for a relatively constant end" (MacNeilage & DeClerk, 1969; according to Daniloff & Hammarberg, 1973, p. 244). MacNeilage (1979, p. 19) summarizes:

> In recent years a number of writers have emphasized the possible role of *auditory* targets in speech gesture control (e.g. Nooteboom 1970; Ladefoged *et al.*, 1972). Informal evidence for the necessity of auditory targets in some sense of the term is quite conclusive. The auditory information provided by our language community is the only source of goals for our acquisition of speech production. A given auditory goal is sometimes achieved in a single subject by more than one spatial configuration of the speech apparatus. For example single intervocalic [p] is produced in English with vocal fold abduction (Lisker et al., 1969). But cluster-initial intervocalic [p] as in "upbringing" is produced in some subjects by vocal fold adduction (glottal stop) (Westbury, 1978). Thus it is the auditory goal that remains invariant in this case at the expense of invariance in spatial configurations.

Earlier I summarized similar arguments that plans for phonetic behavior, i.e. phonological forms, must be defined by conditions on socially observable behavior.

Phonetic plans are abstract conditions on phonetic acts. The psychological reality of such conditions has been demonstrated in various phonetic experiments, in which one has placed subjects in entirely new and abnormal predicaments by obstructing the natural use of articulators. For example, Lindblom has demonstrated that subjects, who are required to produce vowels with bite blocks between their teeth, can immediately produce the correct acoustic effects without training. The fact that the correct formant frequencies are achieved in the first pitch period of the vowel immediately after the insertion of the bite blocks rules out the interaction of acoustic feedback as an explanation, and shows that subjects can immediately adjust the remaining controllable variables of the speech apparatus in order for the whole system to attain the desired acoustic effects defined by (the speakers' knowledge of) the abstract conditions. Similar conclusions can be drawn from an experiment by Riordan (1977) in which subjects were able to produce vowels sounding like ordinary rounded vowels in spite of the fact that lip protrusion was artificially impeded. (The subjects used larynx lowering instead).

Thus, we claim that plans at different levels of specificity do possess psychological reality. This reality must be characterized in terms of conditions on overt phonetic behavior and not in terms of peripheral motor commands. Naturally, we must assume that the speaker's knowledge of the abstract conditions has some kind of central neural counterpart, but there is still a tremendous gap between our knowledge of, on the one hand, neurophysiological processes and, on the other hand, properties of linguistic acts. However, phonetic plans may fulfil a completely decent, explanatory function in a theory of speech performance even in the almost total absence of any detailed neurophysiological knowledge.

16. The perceptual genesis of phonological features

Thus far I have mainly discussed phonological forms in terms of plans for *producing* phonetic behavior. It is now time to counterbalance this bias, and consider the functions of phonological forms within the activities of perceptual exploration and phonetic imagery. First, I will comment on the fact that conditions on phonetical behavior are basically perceptually determined.

We assume that for each expression (word, etc.) in a language L there are certain critical phonetic features the majority of which normally has to be manifested in order for the tokens to be identifiable as occurrences of that expression. Consequently, the speaker has to include these features in his phonetic plan. However, to be able to do that the speaker must have become aware (at some level of awareness) of these features in the phonetic signals produced by other speakers and (at some point of his linguistic development) identified the features as distinctive for the expression in question. Thus, the phonetic features of phonological forms are perceptually determined. We also know that children are in general capable of identifying critical features before they can produce them properly themselves.

Perceptual processes categorize the quasi-continuous flow of physical events into more or less discrete categories. This general principle of perception may explain why phonological forms contain a set of categorical conditions and why we seem to experience speech as built upon a limited number of invariant phonetic types (e.g. syllables, vowels, conconants, etc.) in spite of the fact that phonetic behavior is a quasi-continuous stream. Thus, we assume that language users select a number of distinctive features by perceptual abduction[26] (abstraction) from the flow of phonetic effects.

Phonetic plans not only have a perceptual genesis, they may also be assigned a function

[26] Abduction in the sense of Ch. S. Peirce, Cf. Anttila (1972, p. 196).

in the mature speaker's perceptual activities. It is reasonable to assume that speakers expect to find the critical features of phonological forms in the various phonetic signals and, thus, that they try to reinterpret them into what they hear. We may even assume that phonological forms function as plans (guiding principles) for perceptual activity too.[27] Therefore, I will try to develop Neisser's notion of the perceptual cycle along these lines (Section 17).

However, even if phonological forms contain distinctive features that define the identity of the linguistic items involved, and these features are perceptually determined, it does not follow that these "identity features" are also actually used as "identifying features" or "cues to identification" in the perceptual activities of competent language users. One reason for this resides in the fact that perceptual activities vary with different tasks (cf. Section 17). It is conceivable that perceivers use special tricks in some cases of actual perception and that the cues actually used are not the features actually intended as pertinent by speakers. On the other hand, there is a constant interplay between production and perception, and it seems plausible that if a certain language user interprets certain features as important for the identification of specific linguistic items, then he also tends to produce these phonetic effects when he plays the role of the speaker trying to pronounce the items in question.

17. The perceptual cycle and perceptual expectations

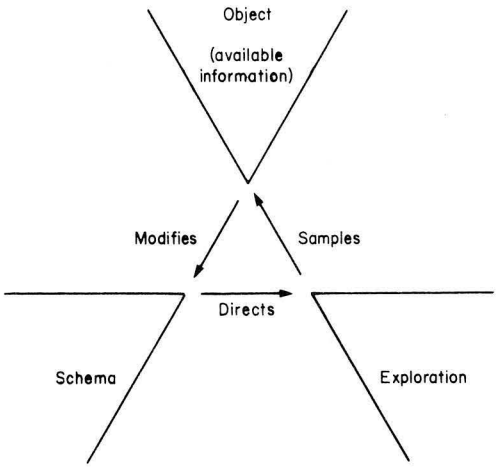
Neisser (1976) develops an interesting theory of perception, in which ideas of J. J. Gibson (e.g. 1966) and J. Piaget are combined. The basic idea of Gibsonian theories of perception is that the perceiving agent actively searches for and identifies certain patterns in the stimuli themselves; he directly picks up the information that the outer world offers him. Piaget emphasizes the perceiver's highly structured strategies or *schemas* (cognitive structures) that enable him to assign structure to the world.[28] Neisser (op cit., pp. 20–21) writes the following about the perceptual cycle:

> In my view, the cognitive structures crucial for vision are the anticipatory schemata that prepare the perceiver to accept certain kinds of information rather than others and thus control the activity of looking. Because we can see only what we know how to look for, it is these schemata (together with the information actually available) that determine what will be perceived. Perception is indeed a constructive process, but what is constructed is not a mental image appearing in consciousness where it is admired by an inner man. At each moment the perceiver is constructing anticipations of certain kinds of information, that enable him to accept it as it becomes available. Often he must actively explore the optic array to make it available, by moving his eyes or his head or his body. These explorations are directed by the anticipatory schemata, which are plans for perceptual action as well as readiness for particular kinds of optical structure. The outcome of the explorations – the information picked up – modifies the original schema. Thus modified, it directs further exploration and becomes ready for more information. This cycle is diagrammed in Fig. 2 [here: (13)/PL].

[27] This follows a suggestion by Öhman (1975*a, b*, pp. 318–9 *et passim*). Öhman also applies the same reasoning to other aspects of linguistic acts. Thus, "the structural meaning of the linguistic act can be understood as a model or as a set of directions, for the listener's perceptual behavior. In other words, the structural meaning of the linguistic act ASKS FOR a (complex) perceptual act. It represents a perceptual plan" (op cit., p. 329).

[28] Neisser also alludes to Bartlett's (1932) notion of *schema*. Similar ideas have recently been suggested by e.g. Norman & Bobrow (1976).

(13)



The perceptual cycle

We may now assume that phonetic plans (and other aspects of expression plans) are somehow involved in the perceptual schemas used in speech perception. Thus, these plans would also constitute conditions on the perceptual activities. These conditions guide the perceiver's perceptual explorations of the phonetic signals ("the available objects"). This means that the perceiver expects to find certain kinds of structures in the incoming phonetic signals, he therefore searches for such structures, sometimes he also manages to recognize (find) them, and his expectations may also lead him to reinterpret structure into signals that are deviant, noisy or defective. However, the phonetic properties actually present constantly force the perceiver to modify and precisize his expectations, i.e. make his perceptual plans more specific or change them in perhaps important respects. Thus, within this perceptual cycle there is an interplay between expectations (schemas, phonetic plans to be applied) and the structure actually discovered.

A perceptual schema is anticipatory, it constitutes a plan for perceptual explorations, in which certain types of information are expected to be found. Sometimes, we become aware of this. For example, if we unexpectedly hear someone speak a foreign language, it usually takes a little time to "tune in" this language; one has to change one's expectations of hearing something that fits the general patterns of one's native language (the general properties of the perceptual schemas of that language), and instead activate the perceptual schemas of the foreign language. Similarly, when we record a certain word in a context where we do not expect it, it may take some extra time to identify this word and thus to understand what was actually said.

Normally perceptual expectations are not very specific. This would mean that we do not activate specific phonetic plans in normal speech perception. However, under special conditions we may be forced to develop very specific perceptual plans, perhaps by iterative perceptual cycling. Suppose, e.g. that we come across a completely unknown word and that we want to find out its phonological structure. Then we must, probably by successive perceptual cycles, determine its phonemic properties (while the rest of its phonetic properties are unnecessary to determine at once, since they follow from rules). Similarly, in certain situations we may be forced to decide which one out of two or more phonetically similar

words is pronounced. In such a case we may use our knowledge of which distinctive properties belong to the phonetic plans of the possibly occurring items also in our perceptual activity. On the other hand, our perceptual expectations may also go beyond phonological forms in phonetic detail, if we are interested in the specifics of particular phonetic manifestations (e.g. idiolectal, dialectal, and phonostylistic features). Thus, perceptual expectations may be general *or* specific depending on the task, i.e. the goal of the perceptual activity.

However, in spite of the facts just mentioned, if we look upon our linguistic experience in global terms, it is important now and then to attend to such features that may distinguish different words, and therefore it is reasonable to assume that our general perceptual habits and expectations are influenced by the phonemic patterns, i.e. (in our theory:) the general properties of phonological forms (= phonetic plans). This assumption also underlies much traditional thinking about speech perception (e.g. Sapir, 1921, pp. 55–6).

18. On perceptual activities in normal speech perception

One can listen to speech in many different ways. Depending on what one is interested in, what the task or goal of the perceptual activity is, one will focus on different aspects and therefore, we assume, use different kinds of perceptual schemas (plans). In normal speech perception, we do not focally attend to the sounds at all. Rather we use fragmentary information at the phonetic level to recognize words, phrases or even "meanings" directly. In fact, it seems impossible to search for "meanings" and at the same time attend to the phonetic properties themselves. As soon as we make ourselves focally aware of the phonetics of the utterance we lose the integrative, higher-level perspective (cf. Polanyi, 1969); afterwards we would have great difficulties to report the semantic content of the text we heard.

The fact that we are only "subsidiarily aware" (Polanyi) of phonetic properties is enough to doubt any assumption that we use full phonetic plans (= phonological forms) in every act of normal speech perception. Rather, we may be assumed to use a complex perceptual schema, in which the goal is to identify words, referents and meanings. Possibly, such a perceptual schema may be seen as a hierarchy in which perceptual activities (plans) of different orders are embedded (cf. Neisser, 1976, p. 113). We aim at grasping "meanings" and spend energy and attention on identifying phrases, words, and phonetic properties only to the extent that it proves necessary for attaining the superordinate goal of finding a reasonable interpretation of the whole communicative act. Thus, perceptual activities may involve multiple embeddings of lower-level acts much like utterance production does (Section 3). Instead of having the mechanical system of a fixed linear sequence of information processing stages in perception (Section 11) we seem to be forced to postulate a number of semi-independent strategies, hierarchies of schemas etc that can be more freely manipulated depending on the goals of the perceptual activities.

There is evidence that word forms, particularly the semantically heavy items, are important primes in speech perception (e.g. Clark & Clark, 1977), i.e. we hear speech in terms of these *meaningful events* (cf. Neisser, 1976, p. 155). It is therefore even more motivated to regard word forms as primary units in phonology (cf. Section 5, where it was argued that phonetic plans are word-form-invariant). Certainly, we do not hear "sounds" (segments) (let alone "phonemes", cf. Section 19), although we can be trained in segmenting speech, especially under the influence of becoming literate. It is therefore hardly surprising that we are so bad at identifying sequences of meaningless sounds even if these acoustic stimuli are much longer than "segments" in normal meaningful speech (cf. Warren *et al.*, 1969). What may seem surprising from our present point-of-view, however, is the stubbornness with which linguists and phoneticians have clung to the view that normal speech perception does

involve a stage of identifying the phonemic segments of the signals perceived. (For example, Liberman *et al.* (1967, p. 452) talk about "the fact (sic!) that we do perceive phonemes").

19. Phonological forms and percepts

The phonological form of a word is not something that can be heard. Phonological forms pertain to perceptual *processes* rather than their results (percepts), and we do not perceive the processes while we are using them to perceive something else. What we hear is the actual word spoken. Furthermore, in normal speech perception we are not focally aware of its sound gestalt at all, and even if we adopt the particular attitude of listening to the sound pattern as such, we cannot hear the phonemic features alone. This is so because (a) we cannot concentrate on several parts of a word at one time, and (b) whatever spot we happen to concentrate on, our concentration will make us detect more than just the phonemic features.

The percept, according to the type of theory proposed here, is the phonetic "object" (event) existing in the signal (behavior). The percept is "out there"; it is not a mental picture or so. Although perception involves categorization in terms of types (cf. phonological forms), what we perceive is the specific token, not the type (category). Percepts are therefore neither "subjective", mental entities nor "objective", directly or automatically extractable physical properties, but they are the outer objective reality viewed (or heard) under a certain point-of-vew (perceived-as-something); "their material support *is* (or, rather, is *often*/PL) available in the world, but is only detected by a specially attuned organism" (Fowler, 1980, p. 121).

20. Phonological forms and images

A phonological form is not a percept, nor is it the result of imagining the sound gestalt of the word form involved, i.e. an image or *Lautvorstellung*. But how should we understand what happens when we imagine the sound gestalt of a word form in the absence of an actually manifested phonetic signal? Notice that this is something we often do when we make ourselves conscious of phonetic properties; we repeat to ourselves, "in our mind", a phonetic gestalt (that may or may not have been actually heard before). Sometimes, we also do this when we carefully plan what we are going to say; we construct an utterance and silently execute it while rehearsing it (Linell, 1978, Section 10).

Öhman (1975*a*, *b*, p. 330) has proposed that when we imagine something, we perform the perceptual operations "vacuously", i.e. in the absence of actual sounds (in the case of speech). The image, the percept of such an act, is "an indirect object of perception" (ibid., p. 331). Neisser (1976, p. 130) argues in somewhat similar terms; he proposes that we may anticipate the structured information we would expect to perceive in a potential environment. Imagining involves performing only the anticipatory phases of perceptual activity, something which is possible once schemas have been "detached" (first a child can only manipulate and sense the outer reality present here-and-now, then he can also imagine it). In imagination, we manifest an ability to divide, detach and manipulate our anticipations (p. 133). Neisser (p. 131) contends that images are not copies of earlier percepts but plans for obtaining information from potential environments. This way of putting it is slightly misleading, however, since plans are more abstract than images. We would prefer to say that images are the subjective experiences that result from applying the perceptual operations vacuously. Therefore, in a sense images do not exist; what actually occurs is simply the activities or processes of imagining the thing involved, and these activities may produce an impression of perceiving something "in our inner eye or ear". But the image is imaginary;

there is no mental object to be found anywhere, e.g. no mental or neurophysiological state in the brain corresponding to the image. The philosophical troubles begin, when we start to think of linguistic forms and rules as mental objects or states rather than as conditions on, or patterns of, ongoing activities.

Thus, *Lautvorstellungen* would be explicable in terms of what we would expect to hear if we had produced or come across a certain string of phonetic behavior. Now, I find it interesting to note that when we imagine the sound gestalt of a word form and attend to its phonetic aspects, then we seem to imagine some sort of careful pronunciation. Likewise, when we ask a person to tell what he has just heard, say on a tape recording, he often reports that he heard a more careful pronunciation of a particular word than was actually manifested, i.e. he interprets more structure into what was heard than was actually there to be heard (so-called phonemic restoration). In our theory, this is natural. If there is no longer any matching signal to influence the outcome of the perceptual activity – the whole perceptual cycle cannot be carried out – then the result, the "image", would be heavily dependent on the stereotype inherent in the perceptual schema, i.e. the phonological plan, which pertains to a careful pronunciation (Section 6).

21. Conclusion

I have argued that the phonological form of a given linguistic item may be seen as a perceptually determined, little-detailed structure specification of the most careful pronunciation of that item (i.e. the most careful pronunciation known to the speaker). This "phonetic plan" may function, together with finer-grained plans, in both productive and receptive activities, as a plan for pronouncing the item in question, as a perceptual schema guiding the activity of a listener expecting to hear that item, and in phonetic imagery.

It is interesting and significant that there are several important currents in contemporary psychology, linguistics and phonetics that converge right now to support such a view. Although belonging to traditions of rather different origins and characters – on the one hand philosophical action theories, on the other empirical explanatory theories in psychology and physiology – these currents seem to have important presuppositions and values in common. Thus, we have the theories of pragmatics and communicative acts developed in the Anglo-Saxon world (going back to the work of e.g. Wittgenstein and Austin) (e.g. Allwood, 1976; Bennett, 1976), the Russian tradition of action theories for complex motor behavior, esp speech (Vygotsky, Luria, Leontiev), theories of perception based on the work of Piaget, Gibson and Neisser (e.g. Neisser, 1976), and recent theories of speech production involving auditory targeting as a key concept (MacNeilage and others, cf. Section 14). Quite recently, some American scholars have applied action theory to speech production (Fowler, 1977, 1980; Fowler *et al.*, 1980). Since these scholars refer to the work of the Russian physiologist Bernstein and his theories of co-ordinated movements (cf. Bernstein, 1976), we may perhaps see a tradition via him to Luria and from Luria and Leontiev all the way back to Baudouin de Courtenay, the founder of phonology. Another tradition going back to Baudouin is that of classical European structuralism (Bühler, Trubetzkoy, R. Jakobson) that provides the foundation of the kind of phonological theory on which we build. It is encouraging to find that the kind of theory proposed here is supported by this rich array of scholarly traditions of different origins.

This paper documents my attempts at attaining further understanding of some problems discussed in Linell (1979). I have worked on this particular paper at intervals since early 1978. The present version was completed in May 1980. In the meantime, I have come

across several important works by other scholars that seem to support some of the fundamental ideas propounded here. In particular, I would like to mention Fowler *et al.* (1980).

I have had the privilege of discussing some of the problems raised here at seminars in Copenhagen, Stockholm, Turku and Uppsala. I have also had stimulating private discussions with many people, and some have contributed important comments on earlier versions of the paper. In particular, I wish to thank Jens Allwood, Jan Anward, Björn Lindblom, Martti Nyman, Svend Erik Olsen, Sven Öhman, Ragnar Rommetveit, and Richard Wojcik. I am also indebted to Annika Axelson for the typing.

References

Allwood, J. (1976). *Linguistic communication as action and co-operation*. Göteborg: Department of Linguistics.

Anttila, R. (1972). *An Introduction to Historical and Comparative Linguistics*. New York: Macmillan.

Bailey, Ch. -J. (1979). Invited comments on congress reports by Basbøll, Anderson and Bybee. *Ninth International Congress of Phonetic Sciences*, Copenhagen.

Bailey, Ch. -J. & Maroldt, K. (1979). Über den gerechten Tod des Phonems. *Working Papers in Linguistics* 4, Technische Univ. Berlin. pp. 63–80.

Bartett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Bennett, J. (1976). *Linguistic behaviour*. Cambridge: Cambridge University Press.

Bernstein, G. (1977). *The Co-ordination and Regulation of Movements*. London: Pergamon Press.

Brown, G. (1977). *Listening to Spoken English*. London: Longman p. 70.

Chistovich, L. & Kozhevnikov, V. (1969). Some aspects of the physiological study of speech. In Proctor. L. D. (ed) *Biocybernetics of the Central Nervous System*. Boston: Little, Brown.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clark, H. & Clark, E. (1977). *Psychology and Language. An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.

Coates, R. (1980). Time in phonological representations. *Journal of Phonetics*, 8, 1–20.

Daniloff, R. G. and Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, **1**, 239–248.

Danks, J. (1977). Producing ideas and sentences. In: Rosenberg, Sh. (ed), *Sentence Production*. New York: Erlbaum. pp. 229–258.

Derwing, B. (1973). *Transformational Grammar as a Theory of Language Acquisition*. (Cambridge Studies in Linguistics, 10). Cambridge: Cambridge University Press.

Donegan, P. J. & Stampe, D. (1978). The syllable in phonological and prosodic structure. In: Bell, A. & Hooper, J. (eds) *Syllables and Segments*. Amsterdam: North-Holland.

Donegan, P. J. & Stampe, D. (1979). The Study of Natural Phonology. In: Dinnsen, D. (ed), *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press. pp. 126–173.

Dressler, W. U. (1977). *Grundfragen der Morphonologie*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Fischer-Jørgensen, E. (1975). *Trends in Phonological Theory*. Copenhagen: Akademisk Forlag. p. 22.

Fowler, C. (1977). *Timing control in speech production*. Indiana University Linguistics Club.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113–133.

Fowler, C., Rubin, P., Remez, R. E. & Turvey, M. T. (1980). Implications for Speech Production of a General Theory of Action. In: Butterworth (ed) *Language Production. Vol 1: Speech and Talk*. London: Academic Press. pp. 373–420.

Fromkin, V. (1973). (ed) *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Garrett, M. (1975). The Analysis of Sentence Production. In: Bower, G. (ed) *Psychology of Learning and Motivation, Vol. 9*. New York: Academic Press. 133–177.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis*, **2**, 23–68.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, **4**, 353–363.

Hasegawa, N. (1979). Casual speech vs fast speech. *Papers from the Fifteenth Regional Meeting of Chicago Linguistic Society*. Chicago. pp. 126–137.

Hooper, J. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Katz, J. J. (1964). Mentalism in linguistics. *Language*, **40**, 124–37.

Keller, E. (1979). Planning and execution in speech production. *Montreal Working Papers in Linguistics*, **13**, 34–51.

Kent, R. D. & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, **5**, 115–133.

Korhonen, M. (1969). Die Entwicklung der morphologischen Methode im Lappischen. *Finno-Ugrische Forschungen*, **37**, 203–262.

Ladefoged, P. (1971). The limits of phonology. In: *Form and Substance*. Ed. B. Spang-Thomsen. Copenhagen: Akademisk Forlag. pp. 47–56.

Ladefoged, P., DeClerk, J., Lindau, M. & Papçun, G. A. (1972). An auditory-motor theory of speech production. UCLA *Working Papers in Phonetics* **22**, 48–75.
Langacker, R. (1972). *Fundamentals of Linguistic Analysis*. New York: Harcourt Brace Jovanovich.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* **74**, 431–461.
Lindblom, B. (1963). Spectrographic study of vowel reduction. *JASA,* **35**, 1773–1781.
Lindblom, B. (1979). Some phonetics null hypotheses for a biological theory of language. *Proceedings of the Ninth International Congress of Phonetic Sciences*. Vol. II. Copenhagen: Institute of Phonetics. pp. 33–40.
Lindblom, B., Lubker, J. & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics,* 7, 147–161.
Lindblom, B. & Sundberg, J. (1971). Neurophysiological representation of speech sounds. Paper presented at the *XVth World Congress of Logopedics and Phoniatrics*, Buenos Aires, Argentina.
Linell, P. (1978). On knowing what one is going to say. In: Linnarud, M. & Svartvik, J. (eds) *Kommunikativ Kompetens och Fackspråk*. Lund: Läromedelscentralen. pp. 1–19.
Linell, P. (1979). *Psychological Reality in Phonology*. A theoretical study (Cambridge Studies in Linguistics, 25) Cambridge: Cambridge University Press. p. 54.
Linell, P. (1980). The synchronic validity of linguistic constructs: comments on the treatment of 'psychological reality' in contemporary linguistics. *PILUS 39* (Stockholm: Department of Linguistics). p. 34.
Linell, P. (in preparation) The written language bias in modern linguistics.
Lisker, L., Abramson, A., Cooper, F. & Schvey, M. H. (1969). Transillumination of the glottis in running speech, *JASA*, **45**, 1544–1546.
Lubker, J. *et al.* (1980). Symposium No. 6: Motor control of speech gestures. In: *Proceedings of the Ninth International Congress of Phonetic Sciences*. Vol. III. Copenhagen: Institute of Phonetics. pp. 267–282.
MacNeilage, P. (1970). Motor control of serial ordering of speech. *Psychological Review,* 77, 182–196.
MacNeilage, P. (1979). Speech Production. In *Ninth International Congress of Phonetic Sciences 1979, Proceedings, Vol. 1*. Copenhagen. pp. 11–39.
MacNeilage, P. & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *JASA,* **45**, 1217–1233.
Miller, G. A., Galanter, E. & Pribram, K. (1960). *Plans and the Structure of Behavior.* New York: Holt, Rinehart & Winston.
Morin, Y-Ch. (1974). Règles phonétiques ou règles phonologiques. *Montreal Working Papers in Linguistics,* **1**, 45–47.
Neisser, U. (1976). *Cognition and Reality*. San Francisco: W. H. Freeman.
Nooteboom, S. (1970). The target theory of speech production. *IPO Annual Progress Report,* **5**, 51–55.
Norman, D. & Bobrow, D. (1976). On the role of active memory processes in perception and cognition. In Cofer, Ch. (ed) *The Structure of Human Memory*. San Francisco: W H Freeman. pp. 114–132.
Öhman, S. (1975*a*). Modelling linguistic interaction in terms of lexical operators. In Dahlstedt, K. -H. (ed) *The Nordic Languages and Modern Linguistics* 2. Stockholm: Almqvist & Wiksell. pp. 304–346.
Öhman, S. (1975*b*). What is it that we perceive when we perceive speech? In Cohen, A. & Nooteboom, S. (eds) *Structure and Process in Speech Perception*. New York: Springer-Verlag. pp. 36–48.
Olsen, S. -E. (1979). On the information-processing paradigm in cognitive psychology and psycholinguistics. In Pettersson, Th. (ed) *Papers from the Fifth Scandinavian Conference of Linguistics*. Part II: General Sessions. (Acta Universitatis Lundensis. Sectio I Theologica Juridica Humaniora 31). Stockholm: Almqvist & Wiksell. pp. 93–108.
Olson, D. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review,* **47**, 257–281.
Polanyi, M. (1969). *Knowing and Being*. Essays by M. Polanyi edited by M. Grene. Chicago: The University of Chicago Press.
Riordan, C. J. (1977). Control of vocal-tract length in speech. *JASA,* **62**, 998–1002.
Rosch, E. (1975). Human categorization. In Warren, E. (ed) *Advances in Crosscultural Psychology*. London: Academic Press. pp. 1–49.
Ryle, G. (1949). *The Concept of Mind*. Harmondsworth: Penguin (1970).
Sapir, E. (1921). *Language*. New York: Harcourt, Brace and World.
Semiloff(-Zelasko), H. (1973). Syncope and pseudo-syncope. *Papers from the Ninth Regional Meeting of Chicago Linguistic Society*. Chicago. pp. 603–614.
Sperling, L. (1977). *Phenomenology and the social world*. The philosophy of Merleau-Ponty and its relation to the social sciences. London: Routledge & Kegan Paul.
Stampe, D. (1979). *A Dissertation on Natural Phonology*. New York: Garland.
Steiner, G. (1975). *After Babel: Aspects of Language and Translation*. Oxford: Oxford University Press.
Studdert-Kennedy, M. (1976). Speech perception. In Lass, N. (ed) *Contemporary Issues in Experimental Phonetics*. New York: Academic Press. pp. 243–93.
Tatham, M. (1969). Classifying allophones. *Occasional Papers* 3, University of Essex (Language Center).

Trubetzkoy, N. S. (1958). *Grundzüge der Phonologie*. (third edition) Göttingen: Vandenhoek & Ruprecht.
Warren, R. M., Obusek, C. J., Farmer, R. M. & Warren, R. P. (1969). Auditory sequence; confusion of patterns other than speech or music. *Science,* **164**, 586–7.
Westbury, J. (1978). Aspects of the temporal control of voicing in consonant clusters in English. Unpublished Ph.D. thesis. Univ. of Texas at Austin.
Wojcik, R. (1979). The phoneme in natural phonology. In: *The Elements: A Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society. pp. 273–284.