Can the Models of Evolutionary Biology be Applied to Phonetic Problems?

Björn Lindblom
Stockholm, Sweden

1. A Biological Style of Inquiry

By way of introduction I would like to make two points. First, ‘How do we characterize a biological style of inquiry?’ From among the several possible ways of answering such questions I shall choose a formulation that I found in a paper by one of the leading figures in the development of modern biology, Dobzhansky (1965) who says that, confronting any phenomenon in living organisms, the Darwinian biologist has to ask three kinds of questions: The question of (i) mechanism: ‘How does it work?’; the question of (ii) function: ‘What does it do for the organism?’; and the question of (iii) origin: ‘How did it get that way?’ (both for ontogeny and for phylogeny).

The point is here that the ideal biologist envisioned by Dobzhansky uses an interactive strategy. He uses the three viewpoints in parallel. He asks both HOW- and WHY-questions to elucidate a given problem.

Suppose we apply this thinking to a phonetic problem. Take the problem of speech units. Let us examine the available experimental evidence on the production of speech and moreover let us suppose that we are unable to find any facts seriously contradicting the idea that speaking involves the conversion of discrete psychological units into continuous physical signals. If we were to apply Dobzhansky’s program to that particular problem it would not be sufficient to provide a description – no matter how detailed – of what human speakers actually do when they carry out the transformation from the discrete to the continuous. Our account must also address the other two issues, viz. the question of purpose and the question of origin of the proposed mechanism.

It appears clear that this three-criterion method is a powerful one in that it imposes rather severe constraints on the class of possible accounts that we might come up with for any given phenomenon. In other words, it could in principle help us choose between competing theories. Clearly, this is a valuable aspect that should contribute towards making a biological approach interesting to us.

For the purpose of our discussion we also need to define what we are going to mean by biological explanation. How do biologists deal with the question of function and origin? As a second point of introduction let us briefly review some aspects of modern theories of evolution that are essential to our
theme. According to one widely accepted school of thought – the so-called neo-darwinian 'modern synthesis' (Mayr 1978) – new species have evolved as a result of natural selection operating on the variation of existing lifeforms. This variation accumulates continually and arises from dynamic genetic processes such as mutation and recombination. The genes of those individuals who survive sufficiently long to have offspring are transmitted to new generations. Genetic material not compatible with survival and propagation tends to be filtered out. Thus natural selection acts as a sort of editor testing the environmental fitness of new genetic messages.

Using a terminology from our own field we could conceive of evolution as a source-filter process in which the properties of the 'source' as well as the characteristics of the 'filter' vary in space and time. This 'modern synthesis' thus teaches us that it is as a result of interaction between an extremely rich source of genetic variation and selective environmental filtering that the morphology and behavior of many organisms have become so remarkably well adapted to their environments and often exhibit great adaptability to changing conditions.

Deliberately simplifying let me summarize our review of evolutionary theory and state the basic formula for biological explanations:

\[ \text{BIOLOGICAL FACTS} = f(\text{GENETIC VARIATION} \ast \text{NATURAL SELECTION}) \] (1)

It says: Biological facts are explained in terms of an interaction between genetic variation and natural selection.

This is a deceptively simple but enormously powerful principle. Now if biologists have been successful in accounting for the enormous diversity of life forms by developing genetics and the Darwinian idea of selection why not follow in their footsteps and try to account for the likewise enormous diversity of speech sounds along similar lines? I will come back in a moment with some specific examples. Let me first invite you all to reflect on the analogous claim for phonetics:

\[ \text{SPEECH SOUNDS} = f(\text{PHONETIC VARIATION} \ast \text{SELECTION}) \] (2)

It reads: Phonetically as well as phonologically, speech sounds can be explained as arising from an interaction between phonetic variation and selection mechanism.

I agree with Peter Ladefoged that it is not immediately obvious how the data that he represents (Disner 1983, Linker 1982, Narve 1982) can be explained in any simple way by applying 'biological principles'. However, before we assign to them a secondary role or dismiss them totally we must begin to define them and systematically study their interplay with other factors. That is what I would now like to do by considering, in a preliminary way, the nature of speech units.

2. Speech units, Self-Organization and System-Generated Structure

In his abstract Ladefoged (1983a) states that the 'units of abstract linguistics – things such as phonemes and features – are of little relevance for speakers and listeners'. He regards such units as social, but not as psychological realities. He returns to such ideas in the proceedings paper (1983b). We can put Ladefoged's claim – a classical topic – in a biological context by introducing the notions of self-organizing and system-generated structure. The theory of self-organizing systems is a relatively recent paradigm that aims at formulating general laws governing the spontaneous occurrence of order in nature (Jantsch 1981). It can be demonstrated that, wherever there is interaction between subprocesses, this interaction obeys principles of considerable generality and will inevitably impose structuration e.g. on such diverse things as matter, behavior or information.

To convey to you the concept of self-organization more clearly I need to digress for a moment and discuss a distinction which is well known to all of us, viz. the idea of form and substance. However, I shall illustrate it with some examples from other disciplines. For instance consider the form of snow flakes (and crystal formation in general), the splash of a drop of milk as displayed in an instantaneous photograph. Or a chemical reaction: The gradual development of so-called spiral waves in a shallow dish. The hexagonal shape of bee honeycomb cells. (For lack of space I omit the pictures shown during the oral presentation and replace them here by referring the reader to my sources: D'Arcy Thompson (1961) and Prigogine (1976, 1980)).

I could add many more cases but it is not necessary. They would all exemplify the same thing: the notion of self-organizing system. They also represent phenomena which would be difficult to describe on the basis of an explicit
and clear-cut dichotomy into form on the one hand and substance on the other. Form and substance are inextricably interwoven. And there is no advance specification of FORM. Let me clarify the relevance of these seemingly far-fetched phenomena to linguistics. I shall do so by telling you how termites build their nests. The behavior of these insects has been mathematically analyzed as a self-organizing system by Prigogine (1976) and I am indebted to Michael Turvey (cf. Kugler, Turvey and Shaw, 1982) for bringing this work to my attention.

Termites construct nests that are structured in terms of pillars and arches and that create a sort of 'air-conditioned' environment. The form of these nests appears to arise as a result of a simple local behavioral pattern which is followed by each individual insect: The pillars and arches are formed by deposits of glutinous sand flavored with pheromone, a chemical substance that the animals are sensitive to. Each termite appears to follow a path of increasing pheromone density and deposit when the density starts to decrease. Suppose the termites begin to build on a fairly flat surface. In the beginning the deposits are randomly distributed. A fairly uniform distribution of pheromone is produced. Somewhat later local peaks have begun to appear serving as stimuli for further deposits that gradually grow into pillars and walls by iteration of the same basic stimulus-response process. At points where several such peaks close come, stimulus conditions are particularly likely to generate responses. Deposits made near such maxima of stimulation tend to form arches. As the termites continue their local behavior in this manner the elaborate structure of the nest gradually emerges.

The nest building can be described simply in terms of three rules: To initiate deposit at random! Next time deposit where scent density is maximal! Apply recursively! Note that in this theory there is no explicit mention of the source of the nest. This form is implicit in the local behavior of each individual. Consequently the form-substance dichotomy does not apply.

After this digression let us return to Dobzhansky's three questions and the source-filter model in an attempt to apply them to the topic of speech units. Suppose that we try to shed some light on how the mechanism of converting discrete units into continuous signals operates by investigating also the purpose and the origin of this mechanism. If there are such things as phonemes what purposes do they serve and where do they come from? The ontogeny of phonemic coding seems to be a case that clearly calls for a self-organizing model since children proceed from holistic vocalizations to adult segment-based speech as a result of circumstances that they have no direct or conscious control over.

This is work that I am currently doing in collaboration with Peter MacNei-lage and Michael Studdert-Kennedy. It will be one of the chapters of a forthcoming book of ours on The Biological Bases of Spoken Language. We explore a hypothesis an early version of which is due to Hockett and which suggests that phonemic coding arose in a 'self-organizing' way from an interaction between vocabulary growth and phonetic constraints. As man's conceptual development was dramatically accelerated a solution to the problem of efficient signal generation and reception seems to have been obtained in parallel.

The basic idea underlying a series of computational experiments is shown in Fig. 2. We begin on the left by specifying a number k that represents vocabulary size. We feed this number into a computer program that assigns phonetic shape to these elements in a sequential manner and in the presence of certain performance constraints. The selection of phonetic signals is made from a larger inventory representing universal phonetic possibilities. A phonological analysis of the k phonetic signals is then undertaken. By systematically varying the variables of this 'word game' we hope to be able to investigate whether speech-like units could arise from an interaction between vocabulary development and production/perception constraints. Note that the backbone of the theory is the 'variation-selection model'.

Suppose we attempt to derive the phonetic properties of a small lexicon of k words in a manner roughly analogous to the termite story. Replacing deposits by syllables we have:

1. Select first syllable at random!
2. Select next syllable so as to optimize a performance constraints criterion.
3. Apply recursively until k syllables have been obtained!

We shall develop this analogy in three steps:

1. First we define 'possible vocalization' or 'possible syllable':

![Figure 2. Schematic diagram showing components of procedure for deriving systems of phonetic signals as a result of interaction between vocabulary growth and phonetic constraints.](image-url)
2. Secondly we define the selection process:
3. And finally we define the performance constraints and the criterion of optimization.

We make the assumption that the syllable is an axiomatically given primitive of our theory. It is a gesture starting from articulatory closure and ending in an open configuration.

In principle there is an infinite number of places of closure and open configurations. For computational reasons it is necessary to quantize these possibilities into a certain number of discrete points. We decided that a sufficiently finely graded sampling of the universal phonetic space would be obtained by using the 7 closure onsets and the 19 open configurations shown in Fig. 3. This yielded a total of $7 \times 19 = 133$ syllables.

By definition each such vocalization is a holistic pattern that would resemble a CV sequence if presented on a spectrogram. Note that this resemblance does not in any way imply that it is analyzed as a sequence of two segments. It should be regarded as a Gestalt trajectory coursing through the articulatory/acoustic/perceptual space!

Now let us proceed to the definition of the selection process. The assignment of phonetic shape to $k$ distinct meanings can be seen as making $k$ choices from a larger inventory of $n$ possibilities, that is from the possibilities that the universal phonetic space makes available. For our present purposes we are considering a fragment only of that space viz. with $n$ equal to $7 \times 19 = 133$ syllable trajectories.

Given these simplifications we have a combinatorial problem, namely:

SELECT $k$ SYLLABLES FROM $n$ POSSIBILITIES IN THE PRESENCE OF CERTAIN PERFORMANCE CONSTRAINTS

We chose the performance constraints according to Fig. 4. It is important to note that optimization takes place at two levels: With respect to individual syllables as well as with respect to pairs and systems of syllables. In the present simulations we explored the following conditions: Perceptual salience is qualified as extent of syllable trajectory calibrated in auditorily motivated dimensions: To exemplify, [+] comes out as less salient than [j]. Extremeness of articulatory gesture applies both to static configurations and to dynamic events: [+] – closures more extreme than [j]’s. [j] represents a more extensive movement than [+].

Articulatory distinctiveness and perceptual distinctiveness are systems parameters. The articulatory dimension is interpreted as sensory discriminability and is computed in terms of ‘articulatory distance’ as specified by an articulatory model. Perceptual distinctiveness is derived by generalizing results on distance judgements for vowels to holistic syllables. For both of these parameters our metric implies that [ [+]] form a less distinctive system than [b[d][G]].

For a given arbitrary syllable we obtain numbers in the top row. For a given pair of syllables we generate numbers in all four cells. Those four values are combined into a single number whose meaning can be described verbally as perceptual benefit per articulatory cost. The formula is which is applied for each additional syllable is:

$$\sum \sum \frac{1}{L_i j T_{ij}} < \text{THRESHOLD}$$
Note once more that both individual syllables and the whole system are evaluated. Recall now that to simulate the development of the 'lexicon' we applied the formula repeatedly for each new syllable and continued this procedure until a system of $k$ syllables had been obtained. We chose $k$ to be 24, a fraction of the total set. Since this method gives results that depend on the initial syllable we repeated the simulations 133 times each time starting from a new syllable. One way of presenting the results is obtained by answering the following question: For a specific configuration of constraints what is the probability of finding a given syllable in the pooled subsets?

We pool all the subsets and plot the frequency of each syllable in the pooled set. We find that the results deviate markedly from a pattern of completely uniform preferences which is the result we would have expected had the derivations taken place in a completely unbiased fashion, that is without any performance constraints at all.

In Fig. 5 the results have been arranged in the form of a two-dimensional matrix with rows representing onsets and columns endpoints. Syllables that did occur (at least once in 133 runs) have a black cell. Syllables that did not occur at all have empty cells.

It is immediately apparent that certain rows and columns have more than one entry. This means that syllables such as [bu, gu, gu] etc. contrast. Rows and columns with multiple entries contain syllables that keep one segment constant while varying the other. They identify minimal pairs. Since by definition all syllables have distinct meanings we might conclude that according to standard procedures these minimal pairs contain distinct phonemes. The existence of [bu, gu, gu] thus appears to suggest that in these derivations /b, ɡ, ɡ/ come out as separate phonemic segments.

How is that possible? At no point in the derivations have we analyzed the syllables as a sequence of two segments. We have defined our vocalizations as holistic events. Our theory does not invoke 'segment' as an explicit construct. Neither does it use explicit 'features' although the use of [b, ɡ, ɡ] and a certain subset of vowels implies a systematic favoring of certain articulatory properties of 'feature' dimensions.

It should be clear from these considerations that there are neither 'segments' nor 'features' on the generation of these phonetic signals and that it is our preceding linguistic analysis that imputes discrete 'segments' and 'features' to them. Just as 'arches' and 'pillars' are implicit in the behavior of the termite the 'segments' and 'features' represent phonological structure implicitly and non-discretely present in the process of selecting the phonetic system. It is as if the phonetic space becomes 'quantally structured' as phonetic constraints interact with a growing vocabulary.

What do we conclude from these results? Let me extrapolate. Suppose we retain the notion of self-organization and manage to elaborate the theory so that more realistic and language-like phonetic systems can be produced. Will we find that our procedure will eventually derive fully discrete segments and features similar to the ones now postulated by linguistic analysis? Or will our model instead reinforce the notion of implicit psychological realities not in on-line speaker-listener behavior but rather structures indicating that phonemes and features are products of some introspective, metalinguistic ability that we possess as speakers and listeners? Clearly these are questions for future research. For the moment let us conclude that the present results although highly preliminary appear to encourage interest in self-organising models and further search for biological precursors of phonological and phonetic structure.

3. Role of Socio-Cultural Factors

In response to Ladefoged’s remarks on the role of cultural factors and 'the whims of fashion', I would like to make two comments. The description of
the phonetic facts we are considering can be seen as an optimization problem. In accordance with that point of view the phonetics of a given language is the result of optimizing a great many dimensions which interact to yield an overall system value:

**OVERALL SYSTEM VALUE < THRESHOLD**

(5)

Assume furthermore that the observed systems need not exclusively represent *optimal* systems but are simply those systems that are *sufficiently optimal* with respect to the threshold criterion. We then realize that there must be many solutions to a given optimization problem. Thus we conclude that also without social factors the biologically based conditions would give us diversity and non-uniqueness.

My second point is the following:

Suppose we now postulate that the optimized parameter is social and perceptual effect per articulatory cost rather than just perceptual effect per articulatory cost:

\[
S_{ij} \cdot L_{ij} / T_{ij}
\]

(6)

In Fig. 6 we present this idea schematically. Universal phonetic possibilities are discretely represented and compared among themselves three times in the

![Figure 6. Schematic diagram showing a possible extension of model to accommodate also social factors. The rows and columns of the matrices are labeled identically and refer to universal phonetic possibilities. For any given phonetic contrast - that is given cell - the model provides social, perceptual and production-based coefficients which are combined into a single number and stored in the matrix to the right. This matrix forms the basis of system selections.](image)

4. Deductive vs. Axiomatic Theories of Phonology

My next comment concerns Ladefoged's pessimism about our being able to formulate a phonetically based, deductive theory of phonology. To be sure we can have no illusions about the magnitude of such a task. But the existence of difficulties does not convince me that there are easy and acceptable short-cuts.

One reason for insisting on a deductive account is based on the fact that the child can be said to derive its phonology deductively. Here is a comment on language development from Rules and Representations by Chomsky (1980, 66-67): "what we should expect to discover is a system of universal grammar with highly restrictive principles that narrowly constrain the category of attainable grammars, but with parameters that remain open to be fixed by experience. If the system of universal grammar is sufficiently rich, then limited evidence will suffice for the development of rich and complex systems in the mind, and a small change in parameters may lead to what appears to be a radical change in the resulting system. What we should be seeking, then, is a system of unifying principles, that is fairly rich in deductive structure but with parameters to be fixed by experience. Endowed with this system and exposed to limited experience, the mind develops a grammar that consists of a rich and highly articulated system or rules, not grounded in experience in the sense of inductive justification, but only in that experience has fixed the parameters of a complex schematism with a number of options. The resulting systems, then, may vastly transcend experience in their specific properties but yet be radically different from one another, at least on superficial
evolutionary biology and phonetic problems examination: and they may not be comparable point-by-point in general' (itracles ours).

Personally I see a favored role for phonetics within Chomsky's program of universal grammar but I think it is too early for phoneticians to share Ladefoged's and Chomsky's interest in primarily those aspects that must be described in purely formal terms and are said to make language a unique and specialized structure.

It is true that claiming that language is in part an autonomous biological structure should make perfect sense from the biological point of view. After all specialization is in one sense what evolution is all about.

However, our quarrel with the proponents of uniqueness and autonomy views is a methodological one. Again, let us follow the example of professional biologists who seem to prefer accounts of evolutionary changes that play down 'quantum leaps' as much as possible and that manage to interpret changes in terms of a minimum of de novo developments. This is a parsimonious null hypothesis that can be called the continuity or the tinkering principle (Jacob, 1977). Applied to our own field its contents would be:

DERIVE SPEAKEN LANGUAGE FROM NON-LANGUAGE (7)

Claiming that language is special as Ladefoged and Chomsky do prejuidges the issue. For any given phenomenon, it should be preceded by an exhaustive search for preadaptations. Before giving up that search and joining the 'formalist' camp we should make sure that, for example, we have not underestimated the structure-forming power of principles operating in the self-organizing systems subsuming language. Although clearly untrue (gen. speciation) the formulation of Linnaeus remains an efficient null hypothesis of biological inquiry: Natura non facit saltum.

5. On Explanation

Formal and functional approaches are often regarded as incompatible in current debates among phoneticians and phonologists (Andersen, 1981, Ladefoged, 1983b). In biology, this issue of functional explanation seems to have an analogy in the question: 'Is all evolution adaptive?'

Functionalism in linguistics if often based on 'utility' arguments. Since many features of both language use and language structure no doubt lack direct utility it appears advisable to take a dim view of such functional arguments. Given the strong emphasis on adaptation by natural selection the reader may at first find such remarks inconsistent.

My point is this: To be able to put linguistic functionalism on a solid basis we need to learn our biology lessons well and avoid caricatures such as viewing 'each bit of morphology, each function of an organ, each behavior as an adaptation, a product of selection leading to a 'better' organism'. Darwin believed in adaptive and nonadaptive change and pointed to two principles underlying the latter: (1) organisms are integrated systems and adaptive change in one part can lead to non-adaptive modifications of other features;... (2) an organ built under the influence of selection for a specific role may be able, as a consequence of its structure, to perform many other, unselected functions as well.' The current utility, or inutility, of a structure permits no assumption that selection did, or did not, shape it in a direct way. It may have been selected indirectly as a part of a larger system or through a cumulative action of collective subprocesses. (Quotations from an essay on the human brain (Gould, 1980, 59).

Applying this thinking to our own field a lesson for the phoneticians would be that some linguistic phenomena are truly the results of 'adaptive changes' and could thus in principle be explained in functional terms whereas others have arisen nonadaptively and have to be accounted for on a purely formal basis.

In order to arrive at both functional and formal interpretations it would seem that our biologically inspired approach must obey the continuity principle and first lead to exhaustive investigations of all kinds of functional arguments (cf. above). The lesson for the phoneticians would in the light of such reasoning be that the very existence of non-adaptive mechanisms in evolution would not a priori make formal, 'non-adaptive', accounts of linguistic observations legitimate until the search for pre-adaptations had been reasonably thorough.

Ladefoged (1983a) states that 'much of our work as phoneticians is simply to provide good descriptions of linguistic events' and that 'phoneticians must be able to document' language 'differences without expecting to explain them'.

His comments are reminiscent of an often quoted remark by Martin Joos (1958; 96) who wrote: 'Trubetzkoy phonology tried to explain everything from articulatory acoustics and a minimum set of phonological laws taken as essentially valid for all languages alike, flatly contradicting the American (Boas) tradition that languages could differ from each other without limit and in unpredictable ways, and offering too much of a phonological explanation where a sober taxonomy would serve as well.'

To be sure there will be limits to what we may be able to explain but in my opinion we are still far from having reached the end of our resources. We have a long way to go before phonetics ceases to be an interdisciplinary field and achieves a synthesis of its subfields. That development is under way as evident from this conference and will no doubt bring us closer to some of the long-term explanatory goals.

Secondly, as an inhabitant of a sometimes dark and cold country, let me point to an American tradition different from the one that Joos talks about, viz. the power of positive thinking. Believing or not believing in long-term explanation is clearly going to make a big difference for how we choose our short-term goals.

Thirdly, the issue of explanation is closely connected with the practical use
Evolutionary Biology and Phonetic Problems

of phonetics. All over the world to-day scientists including phoneticians feel an increasing pressure of having to produce practically useful results. One way to meet this legitimate challenge is to work for better theories and better explanations as a basis for improved applications. When a speech therapist, an engineer or a language teacher turn their backs on theory we should interpret such behavior more as an indication of the quality of our present explanations than as a confirmation of theory being in principle irrelevant.

The issue of explanation need definitely not be an academic ivory tower pastime.

Acknowledgements

I am particularly indebted to Peter MacNeilage and Michael Studdert-Kennedy. Much of the work reviewed here is the result of our current collaboration.

Special thanks are due also to Jack Fromkin, Bob Harms and Jim Lubker for instructive and inspiring discussions.

1. Current evolutionary theory is a dynamic field full of controversies such as the sociobiology issue (Rose 1982a, 1982b) and the criticism leveled at the 'modern synthesis' by Stephen Jay Gould and others (Gould 1982). For a detailed review of those questions I must refer the reader to the bibliography.

2. The different distributions of vowel qualities in Yoruba and Italian might conceivably be correlated with other factors in the two phonologies e.g. functional load patterns of contrast. An 'uneven distribution' of vowel qualities as in Yoruba does not immediately invalidate all possible 'biological explanations'.

3. For a recent discussion of the role of biological and cultural factors in language change and evolution, see Wang (1982a, 1982b).

References