『인문언어』(Lingua Humanitatis) 5 (2003): 221-264

Phonetics and Language as a formal System

Robert F. Port and Adam P. Leary
(Indiana University, Bloomington)

Introduction

The most fundamental assumption about language in the academic world for the past century or so has been that *'language is a kind of knowledge'*. As pointed out by Chomsky, this idea underlies most thinking about language for several hundred years. Chomsky's achievement was to follow out the consequences of the specific idea that linguistic knowledge might be *symbolic* in form, that is, that linguistic knowledge could be fully expressed using some formal algebra of linguistic symbol tokens. The goal of linguistic research, on Chomsky's view, should be to discover the formal algebra that is employed in linguistic activity (Chomsky, 1965). The advantage of the symbolic-knowledge assumption is that it permits exploitation of all the power of discrete mathematics to model linguistic knowledge. As with any formal system, description of the system requires spelling out some apriori set of symbol types from which complex representations could be constructed. Thus, for arithmetic one must postulate the

【Keywords】 competence, performance, formal symbol structure, linguistic timing, human brain, a priori phonetics, nondiscreteness in phonology

integers, for propositional logic *p* and *q* and, for linguistics, Chomsky proposed S, NP, [Voice], and so on.

However, There is an awkward consequence of the assumption that language is a form of symbolic knowledge -- that formal symbols are static. Just as someone's knowledge of, say, who the President of the U.S. is seems to be a description of the state of the person's memory system at some point in time, similarly, one might suppose that linguistic knowledge is describable at a static structure at some specific point in time. The first reason this is awkward is that whenever a speaker actually uses language, either by talking or by listening and paying attention to speech, they are actually doing so as an activity in continuous time. But even more awkward is that many details of the temporal patterning of speech turn out to be critical to the proper specification of the linguistic structure. The most important aspect of Chomsky's differentiation of Competence from Performance is that competence is static knowledge (even if described using a generative grammar) and performance is an act or event in time. This distinction is also supposed to distinguish what is characteristic of a specific language (since, as we said, all of language is knowledge) from what it invariant across human speakers. So, the third reason the language-is-knowledge assumption is awkward, is that the timeless-and-static vs. temporal-and-active distinction does not apparently line up wish properties-of-language vs. properties-of-speaker (as will be shown below). Linguistics has followed Chomsky's insights but doing so has forced linguistic thinking to rule out of the field many phenomena that are relevant for understanding language. The goal of this paper is to, first, outline just why the assumptions of modern linguistic theory create a serious problem for understanding time, and, second, what

some of the specific linguistic phenomena are that create difficulties fur linguistic theory.

Sectien 1 : Time and the Formality of Language

No one would deny that speech is produced in time, that is, that the sentences, words, consonants and vowels of human language are always extended in time *when they are uttered.* Still, if language is viewed as form of symbolic knowledge representation, then one must conclude that temporal extension is not an intrinsic property of language and that the temporal patterns of language (other than those representable in terms of the serial order of symbols) will not be relevant or revealing about language itself. Linguistics assumes that the temporal layout of speech is a property that is imposed on language from the outside at the point where the serially ordered symbol structures of the language (that is, of linguistic knowledge) are performed by the human body. It is segmental, discrete transcriptions that represent "the phonetic capabilities of man" (Chomsky and Halle, 1968, p.295). Serial lines of printed text are assumed to be, in many essential respects, good models of actual cognitive representations. The cognitve form of language has serially ordered, discrete words composed from a small inventory of meaningless sound-related segments, just like a printed page.

These cognitive symbol strings may be 'implemented' in time by the linguistic 'performance' system if and when linguistic structures happen to be spoken. One might say that ***speech is language as filtered by (or distorted by) the performance system*** – the system that maps language

into speech. From the traditional linguistic point of view, speech performance is thus derivative and is merely one possible 'output mode'--just one of several ways (along with writing) to get language from the mind out into the body and the world. Speech just happens to impose time on a fundamentally nontemporal structure.

This point of view seems to be fundamental in all 20th century structuralist views of language (de Saussure, 1916; Bloomfield, 1926; Hockett, 1954) but most explicitly so of the generative paradigm (Chomsky, 1965; Chomsky and Halle, 1968). On one hand, here is a formal world, an aspect of the mind, the Competence World, where the serial order of hierarchies of timeless symbols provide the data structures of natural language. Formal operations apply to these data structures just as they apply in a derivation in formal logic or mathematics. And just as in the formal structures of logic, mathematics and computer hardware. Complex structures are assumed to have building blocks from which they are built. Given the formal nature of the structures involved, any time that might happen to be required for the operations on structures to take place is merely epiphenomenal and is not directly relevant to the formal operations themselves. And on the other hand, there is a physiological world of brains and bodies living in continuous time. From this traditional perspective, the time-free structures of language are "implemented" and processed in time (see Scheutz, 1999 for more on the notion of implementation). Such implementation processes may hold some interest, but they are in no way the natural home of human language. Certainly, linguistics can easily afford to ignore performance issues.

We believe this point of view Is deeply mistaken. Although there are many reasons why (see van Gelder and Port, 1995; Thelen and Smith, 1994; Clark, 1997), we will discuss just a few. The first is that the dichotomy of Competence and Performance creates a gulf that, once postulated. turns out to be impossible to span using the methods of empirical science. This is surely one reason why linguists frequently consider disciplines outside linguistics irrelevant -- experimental psychology, neuroscience and experimental phonetics -- sinoe these time-dependent fields can have no direct impact on language as a pure symbol system. And correspondingly, this is why scientists from other disciplines frequently have difficulty understanding what linguists are doing. Disciplines like neuroscience and much of cognitive psychology lie across the formalism gulf from linguistics. Thus far, no satisfactory way to bridge this conceptual gap has been found. If one assumes that cognitive and linguistic events do *not* take place in space and time and that real physical events *do,* there is no obvious way (other than a mere implementational hack like discrete sampling) to get them together.

Formal Symbolic Systems

To appreciate this problem, it is helpful to review some of the essential properties of a format system. Although linguists assume the symbolic nature of language - at all levels from phonetic segments, to phonological units, morphemes, words, phrases and sentences - less attention has been paid to exactly what properties a symbol token must exhibit in order for the computational system to work as intended. In western science, it seems symbols are employed in three distinct

domains: *for doing mathematical reasoning* (e.g., math, logic, etc), *in software* (e.g., programming in Lisp) and *as a theory of cognition* (e.g., Chomsky; Newell and Simon, 1972. Thus, for various types of mathematioal reasoning, logic uses tokens like *p* and *q* and arithmetic might use *integers.* In formal reasoning (like doing logical proofs or long division, or writing a computer program, etc), operations are performed on symbolic structures as executed by trained human thinkers. Throughout training and professional practice, steps in a formal reasoning process are typically supported by body-external props. That is, formal reasoning requiring more than a step or two depends on external 'scaffoldling' (see Clark, 1997) such as by writing physical symbol tokens on paper (and, very recently, by using the support of programs running on a computer). In computer hardware, formal methods are automated by the use of symbol tokens coded into bits in a digital computer. The third domain for symbolic theories lies in a particular view of various cognitive operations involved in human language and human reasoning (Chomsky, 1965; Fodor,1975; Fodor and Pylyshyn, 1988). The symbol tokens in language (and probably in general cognition) are the words and phonological structure of some language.

As clarified by Haugeland (1985), in order to function as advertised, the ***symbol tokens must be digital***, that is, discretely distinct from each other and recognizable more or less infallibly by the available computational or cognitive equipment. This is an absolute requirement in order for the computational mechanism to manipulate the symbols during processing without error. The atomic units from which all linguistic structures are constructed must have physical discreteness

because it is only their physical form that determines what operations apply to them (Foder & Pylyshyn, 1988). In a computer, the reading and writing operations make errors only once in many trillions of cycles. This is equivalent to assuming, at the level of syntax, that an NP can be infallibly distinguished from a VP. For the program-executing device, units should either be the same or else distinct.

Second, all symbols and symbol structures ***must be either apriori or composed from apriori components.*** Some set of apriori units must be available at the time of origin of a symbolic system from which all further symbol structures are constructed. In the case of logic or mathematics, an initial set of specific units is simply postulated e.g., "Let there be the integers (or proposition *p* or points and lines, etc.)" In computing, physical bitstring patterns (that is, voltage patterns) cause particular operations to occur in discrete time, but the units and the primitive operations were engineered into the hardware itself, and are thus obviously apriori from the point of view of the programmer. According to Chomsky and Halle, it is fairly obvious "that there must be a rich system of apriori properties - of essential linguistic universals." This follows from the fact that children acquire language very quickly with no tutoring despite wide differences in intelligence (Chomsky & Halle, 1968, p.4). The child is able to use, e.g., his innate phonetic alphabet to represent words and morphemes spoken by those around him. So a problem for the symbolic modeling of human language is that we don't know what the apriori symbol tokens are. The discovery of the list of innate primitive units is the one of the primary missions of research in modern linguistics (Chomsky, 1965). Most often linguists assume that the initial list of primitives includes at least units

like [Vowel], [±Voiced], [Noun], [Past Tense], [Sentence] and to forth. These atoms support the construction of complex descriptive statements about various languages by the language learner (or by the linguist).

The third property of symbols, although one that Haugeland did not comment on, is that they must be ***static.*** Since symbolic or computational models function only in discrete time, it clearly must be the case that at each relevant time point (that is, e.g., at each tick of the discrete-time clock), all relevant symbolic information be available. For example, if a rule is to apply that converts apical stops into flaps, then there must be some time point at which the features that figure in the rule, [+stop], [+voice], [+apical] etc. are all fully represented and either are holding steady or somehow are constrained to synchronize with each other while the rule applies in a single step of discrete time. Thus, properties in a symbolic system cannot unfold asynchronously or be distributed across continuous time but must, at the relevant clock tick, be sitting there with some discrete symbolic value. (Of course, there is nothing to prevent simulation of continuous time with discrete sampling, but this is not what the computational hypothesis about language claims.)

Finally, it seems clear that the apriori symbol tokens must come from ***a fixed apriori list***. But, importantly, it seems the list must be ***small in size*** relative to the range of phenomena covered by the theory. If new aprioris may be added without limit, the theory becomes adhoc. And if the innate vocabulary of phonetics is too large, then accounting far the rapid acquisition of language will become problematic. For example, if there are not just a frow values of voice-onset time, but a hundred or more, (in order to account, say, for many language-

specific and context-determined differences), then, in general, repetitions of the same word in different contexts or by different speakers will tend not to be transcribed the same. How could one learn a word – if it is represented differently every time one hears it? So the innate phonetic alphabet must be fairly small to keep this problem under control.

Now, how can this kind of symbolic unit exist in a human brain? True format symbols actually assume some rather nonbiological properties. It is one thing for humans to manipulate arithmetic symbols in a deliberative way leaning on the support of paper and pencil so each step can be written down and checked for accuracy, and for computers to employ specialized discrete-time hardware to process symbolic structures. But it is another matter altogether to assume that genuine formal symbol structures are actually processed in a discretized version of real time by human brains. The problem is that if we study language as a facet of actual physical human beings (rather than as a particular instance of an idealized Platonic system), then its processes and its products must have some location and extent in real time and space. After all, this is true of a computer – the purest example of an implemented symbol system. Bitstrings are discrete, with a tiny vocabulary size and exist in a real time and physical space.

Similarly natural language should be accessible to scientific research methods that investigate events in space and time – real events in real time. Even if there is temporally discrete behavior of the human brain (as suggested by oscillations in EEG), clearly the best way to study this phenomenon is by *gathering data in continuous time* – in order to

discover just where temporal discreteness can be observed and to understand how the discrete-time performance is achieved. Assuming there is a sharp apriori divide between language as a *serial-time structure* and speech as a *real-time event*, is a very risky bet. And, in our view, there is now a great deal of evidence that it is simply false.

The second reason for rejecting the view that language is essentially formal is that it seems clear that, from a biological viewpoint, language is fundamentally and essentially a *spoken medium* not a written one. Contrary to the typical practice of phonologists who take discrete phonetic transcriptions as their input data, all written versions of language (whether orthographic or phonetic transcription) are derived from speech by perceptual processes that are still not well understood and which depend on the transcriber's native language to an unknown degree (see Strange, 1995 for a review of many issues; Logan, Lively & Pisoni.19NN). It is especially in written language where the symbol-like characteristics - like near - discreteness, timelessness and closed inventories of symbol tokens - are most pronounced. Yet all written language is based upon historically recent, culture-dependent writing methods dating back only a few thousand years, using cognitive processes that may be themselves partly dependent on literacy, logic and mathematical generalization. Even today fewer than half the human population is literate and only a minute fraction could appreciate the meaning of a diagram of a sentence or a syllable even if a linguist spent some time explaining these images to them. Why is this so difficult to do?

These seem to us to be real problems for the traditional view of

language as thoroughly formal – problems that cannot simply be brushed off with assertions that we don't yet know much about how the brain works. After all, if every human utterance is built from discrete building blocks (analogous to bitstrings), that is, if linguistic expressions are always discrete structures assembled from linguistic atoms (the way the words on this page are composed from an inventory of letters), then why isn't it always equally transparent to speakers as well as to linguists what the data structures and atoms actually are for any utterance in any language? Back in the 1940s and 50s linguists wrestled with this very issue by tweaking their definitions of *phoneme* and *morpheme,* etc. to accord with the data (e.g., Harris, 1942, Hackett, 1947). Chomsky and Halle swept these issues aside by dismissing surface-structure notions of phoneme and morpheme as missing the underlying symbolic level that could be revealed through understanding the operation of phonological and syntactic transformational rules. They were confident that symbolic simplicity lay just a little deeper. But in the past 35 years, the transformations between Surface and Deep have not become clear but rather increasingly obscure. So the original question needs to be asked again:

Is language constructed entirely from a finite set of apriori discrete symbol types?

If this property can be relied on, then when difficult cases of linguistic description arise, linguists can be confident that whatever the correct description is, it will be discrete, static and constructed from a short list of apriori atoms. That is, linguists can assume that *language really is formal*, and for the case of phonology, that the phonology structures of each language are formal.

The proposal that *the sound contrasts of a language are discrete* from each other is supported by various kinds of empirical facts that bear reviewing:

1. *Minimal sets of lexical items in the dictionary.* Looking at English vowels, we find word sets like *beat, bit, bet, bat and seal, sill, sell, Sal and reefer, rift, left, laughter*, etc. Similarly for consonants. cf. *bad, pad, Bill, pill, black, plaque, Libby, lippy.* All languages have many such tables of minimally distinct sets of words. The key observation is that there seem to be no cases exploiting categories between these vowels and consonants. The ubiquitous observance of such word sets strongly suggests that all languages employ some discrete set of Vs and Cs for "spelling" lexical items.
2. *Introspection.* When we listen to someone making a vowel glissando from, say, [i] to [æ], the vowel seems to perceptually jump from [i] to [I] to [ɛ] to [æ].
3. *Categorical perception experiments.* Experiments on the identification of vowels and consonants varying along acoustic-phonetic variables show that native-speaking listeners have sharp category boundaries between which they exhibit greatly reduced ability to discriminate stimulus differences (Liberman, 1967).
4. *Experimental phonetic results: within-language and cross-language.* Looking just within any language, some experimental production data show relatively discrete patterns. For example, looking at English, in word initial position, the VOT of multiple tokens of *tip* and *dip* will be largely nonoverlapping. Even looking across languages, the variable of VOT seems to exhibit just three modes (Lisker and Abramson, 1964, Figure 8).
5. *Some sounds appear in many languages.* Another suggestive fact is that some sounds appear to show up in many different languages suggesting they may be drawn from a common inventory. Thus many languages have such sounds as [i, a, u, d, b, n, l, t] etc.

On the other hand, every linguist *also* knows that very often it is *not*

obvious what phonological units there are in a stretch of speech or even how many there are. And many languages or groups of related languages have sounds that are unique and are observed nowhere outside their group, depending on how much phonetic detail is examined. Of course, how any transcription is done will depend on one's theoretical assumptions about phonemes or other phonological units. To take one example, is an English syllable like *'chive'* made of 3 sound units (CVC), or 4 (e.g., CCVC) or 5 (as CCVVC)? Both the initial consonant and the medial vowel each have either two parts or a gliding motion both acousticauy and articulatorily. Or, for another example, is the vowel in *beer* the same as the vowel in *bead* or *bid*? Problem cases like these are found everywhere in every language -- as everyone knows who has ever tried to write a fragment of a phonology.

In these typical cases the phonological analyses are not obvious at all. Instead, linguists must bring theoretical principles to bear in order to justify one analysis over others. And the most important and fundamental assumption currently appealed to for making such analytical decisions seem to be the Symbolic Phonology hypothesis (SP) and the associated corollaries, all of which are explicitly endorsed in Chomsky and Halle's *Sound Patters of English* :

1. ***The Symbolic Phonology Hypothesis (SP):*** *Every utterance (in every language) is constructed exhaustively of phonological symbols that are either synchronous (that is, in the same segment) or* serially ordered. *Complex units are composed from simpler units in a hierarchy of levels. Words are constructed from serial segment strings and segments are constructed from synchronous phonetic features.*

2. *Corollary 1.* ***Atomic Inventory:*** *Languages construct their phonological symbols from a subset of the universal set of discrete phonetic features. These features and segmental organization are innate and available to support language learning in infancy.*
3. *Corollary 2.* ***Segmental Organization:*** *Speech sounds are organized discretely in time into independent phonetic segments that, in principle, may occur in any order.*

All significant differences within the sound systems of any language as well as all differences between languages are believed to be representable in this alphabet of innately provided segmental symbols. Languages, it is claimed, can never differ in the continuous-valued implementation of these minimal symbols. (If they could, then languages would not be entirely symbolic.) For spoken language, the segmental distinctive feature vectors guarantee the discreteness of all ocher linguistic units spelled from them. The discreteness assumption underlies every question the linguistic analyst faces: "One segment or two?", "Do I need another rule here or just a modification of one I already have?" or "Is this phenomenon just a performance habit that does not reflect a formal property of the language?" and so on. Any time a problem arises, the assumption is that if you cannot discard the phenomenon from language altogether, then a symbolic description will be needed.

We will present evidence that words and other apparent linguistic units are sometimes merely nondiscretely different from each other (unlike printed letters). It is possible that linguistic units like words and phonemes are not always timeless static objects, but turn out to be necessarily, essentially, temporal. By this we mean that they are

defined in terms of nonsegmental properties (such as duration) distributed widely across a syllable. If there exist any linguistic structures in any language that are *essentially temporal* (as opposed to merely implemented temporally), or if a case of genuine category nondiscreteness exists, then the bold Symbolic Phonology assumption would be seriously compromised.

In following sections, we will present evidence of at least one example of each for English, reviewing some of the evidence for the nondiscreteness of certain phonological patterns and also demonstrating a pattern that is 'essentially temporal'. These results violate the Symbolic Phonology hypothesis and support the view that phonological systems may exhibit some degree of discreteness (that is, some symbol-like properties), but also have many properties that are quite unlike formal systems.

Section 2 : Some Facts about Linguistic Timing

Research on speech production and perception has shown from the earliest era in the mid-1950s that manipulation of aspects of speech timing could influence listeners' perceptual judgments. Thus, vowel duration may influence judgments of vowel Length and consonant Voicing in many languages and voice-onset time influences judgments of Voicing and Tensity - to mention just a few examples (see summaries by Lehiste, 1970, and Klatt 1976). So linguistic theorists had to address the problem of the discrepancy between symbolic phonetic transcriptions and a real-time description of speech. Chomtky and Halle dealt with the problem by postulating universal *'implementation*

rules' to convert serially ordered segmental feature vectors into continuous-time speech gestures. Later Halle and Stevens (1971) proposed some hypothetical implementation rules that would interpreted, for example, a static (synchronized) feature of glottal tension as causing a delay by a certain number of *ms* in the voice-onset time after the release of a stop. So the temporal effect of long-lag VOT was interpreted as epiphenomenal due to a change in a (synchronous) feature value (cf. Lisker and Abramson, 1971, whose argument is similar to that of this essay).

Notice that this solution rests on an important claim about the phonetic implementation that may prove vulnerable. The Halle-S tevens-Chomsky account of speech timing is tenable relative to their theory only if the phonetic implementation processes are universal. For only if the implementation of discrete phonetic symbols works the same for all languages could it be true that utterance are composed entirely of symbols and differ from each other (linguistically) only in symbol-sized steps. The phonology is supposed to specify the language-specific properties of speech, while the phonetic inventory and its implementation is universal. This must be true if the phonetic space is to include all "the phonetic capabilities of man" (Chomsky and Halle, 1968). The following sections present some evidence that appears incompatible with the now traditional story of a universal discrete phonetic inventory.

English and German voicing. Data gathered over the past 30 years make it fairly clear that what distinguishes some pairs of words in English is an *intrinsically temporal property*. English and German

seem to offer a case where two sound classes differ from each other in a particular *durational ratio* between some adjacent acoustic (or articulatory) segment. English has a contrast among stops and fricatives between those transcribed with /b, d, g, z, .../ and those transcribed with /p, t, k, s, .../ differing in [±voicing] or [±tensity]. In English, pairs of words like *'lab-lap, build-built'* and *'rabid-rapid'* contrast in this feature, as do German *Bunde-bunte* (club-Plur, colorful-nom., sing.)

One characteristic of this contrast in both languages is that it depends significantty on a pattern of relative timing to maintain the distinction. If two segment types differ from each other in duration, one might argue that this results from a static feature that has unavoidable temporal effects. But if specification of the feature requires comparing the durations of two or more segmental intervals, then the claim that this is achieved by implementing neighboring segments in a way that preserves their durational ratio begins to strain credibility. For words with syllable-final or post-stress voiceless consonants, like English *'lap, rapid, lumper'*, the preceding stressed vowel (and any nasal) is shorter while the stop closure is longer in the /p/ words relative to the corresponding words with /b/ (e.g., *'lab, rabid, lumber'*) (Lehiste & Peterson, 1960; Lisker, 1985; Port, 1981).[1] That is, the vowel duration to stop duration ratio changes from values around unity for voiceless obstruents like /p/ and /s/ and values of 2 to 3 for voiced obstruents like /b/ and /z/.

---

1) The other main cue for this feature is glottal oscillations during the closure - but in English stops without glottal oscillations still sound voiced if the closure duration is short enough relative to the preceeding vowel.)

Of course, since speakers typically talk at different speaking rates, the absolute durations of the segments are highly variable when measured in ms. For example, Port (1981) had subjects produce minimal word sets like *dig, digger, diggerly,* and *Dick, dicker, dickerly*. The stressed vowel /I/ became shorter as additional syllables were added, but the ratio of vowel duration to stop closure duration remained nearly constant in all the words. The ratios did change, for example, between the wordset above and *deeg, deeger, deegerly,* etc., since the vowel durations are affected by the vowel change (from /I/ to /ʊ/) while the stop closure are not. Clearly absolute durational values (e.g., in milliseconds) cannot be employed to specify the voicing information, since in that case listeners would both produce and perceive more /p/s and /s/s at slow rates and more /b/s and /z/s at faster rates. But the *ratio of V to C* tends to be relatively invariant over many changes in context.

A second kind of evidence for the significance of this durational ratio is that in perceptual experiments with edited natural speech or synthetically constructed speech confirm that it is the relative durations that determine judgments between minimal pairs like */lab-lap/* and */rabid-rapid/* whenever other cues to the voicing feature are ambiguous (that is, in particular, when the consonant closure does not have glottal pulsing) (e.g., Lisker, 1985; Port & Dalby, 1982), Port (1981) called this relationship "V/C ratio". The relative duration of a vowel to the following obstruent constriction duration (or its inverse C/V). This ratio is relatively (though not perfectly) invariant across changes in speaking rate, syllable stress and segmental context as shown in Figure 1 (Port, 1979; Port & Dalby, 1982).

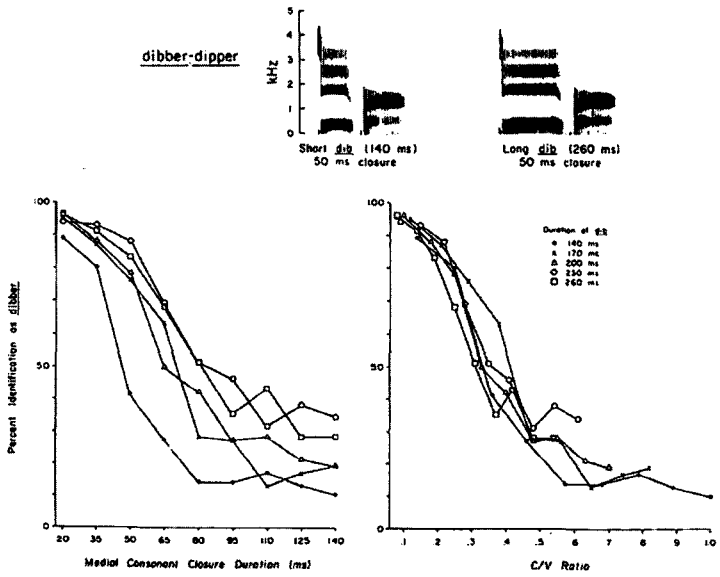

Figure 1. Illustration of some stimuli and results from Experiment 1 of Port & Dalby's (1982) study on consonant/vowel ratio as a cue for voicing in English. The top panel shows sound spectrograms of some synthetic stlmuli. These examples show the shortest (140 ms) and longest (260 ms) vowel durations for *dib*. For each vowel duration step, nine different silent medial-stop closure durations were constructed. Subjects heard such stimuli and were asked if they heard *dibber* or *dipper*. The bottom panel provides the results of the forced choice identification as *dibber* or *dipper*. The left bottom panel shows the identlfication scores as a function of medial stop closure duration. For the shortest vowel the C duration at the crossover is about 50 ms and for the longest, over 80 ms. The bottom right panel shows the same data, plotted as function of consonant/vowel ratio. Note the large reduction of variation as all perceptual boundaries (50% ID) for *dibber* vs. *dipper* cluster near a C/V ratio of 0.35.

In several other Germanic languages, similar measurements of speech production timing (Elert, 1964; Port & Mitleb, 1983; Pind,

1995) and perceptual experiments using manipulations of V and C durations have shown similar results -- that listeners pay attention especially to the relative duration of a vowel and the constriction duration of a following obstruent, that is, stop or fricative (Port & Mitleb, 1983; Pind, 1995, Bannert, 1975). In Swedish the long V-short C vs. short V-long C contrast is partly independent of voicing, with minimal pairs like *vit-vitt* [vi:t, vit:] (white-Basis, white-Neuter) and *bred-brett* (broad-Basic, broad-Neuter) (Sigurd, 1965) and Icelandic *baka-bakka* (to bake, burden-Acc) (Pind, 1995). A similar timing pattern was probably a characteristic of the ancient Germanic proto-language of 2 thousand years BP and has been inherited in somewhat different form by most modern Germanic languages.

Could the V/C durational ratio be a temporal universal? One might claim there is some universal nontemporal feature that causes these durational ratios. But this is surely adhoc. It is one thing to say that some static feature causes a delay or lengthening of some segment, but quite another to claim that a timeless feature causes adjustment of the relative duration of a vowel to a following consonant closure.

Temporal Implementation Rules. Even leaving aside these concerns, there are still major difficulties with any rules of temporal implementation that depend on phonetic context. Since the rules are static, they must specify a duration as some kind of number, that is, as something static that will be interpreted as duration by the performance system. Let's assume for the moment that implementation rules supply an inherent duration *in* ms for each segment type, e.g., 45 *ms* for a [b] closure and 60 *ms* for a [p] closure. Then a context

implementation rule adjusts the duration of the preceding vowel to be longer before a [b] (or shorter before [p]). The result of such rules would be a target duration in *ms* for both the vowel and consonant closure (see Klatt, 1976; Port, 1981; van Santen, 1996 for temporal implementation schemes of this general form).

The first problem here is the issue of what the use of these target durations in milliseconds might be. Who or what will be able to use these numbers to actually achieve a target duration of *N ms* for some segment?

There is no existing model for vertebrate motor control that could employ such specifications. We need a new theory of motor control to make use of these "specs" to generate speech gestures with a specifed duration (see Fowler, Rubin Remez and Turvey, 1980; Port, Cummins, & McAuley, 1995).[2] Second, durations in milliseconds seem fundamentally misguided since speakers talk at a range of rates. So for this reason alone, it seems that it should be *relative durations* that any rules compute, not absolute durations (see Port, Cummins, & McAuley, 1995). Third, since in this model durations are specified one segment at a time, longer intervals (such as intervals between stressed syllables) can get their duration only by adding up the individual segments that comprise them. But such a system has no apparent way to obtain global timing patterns (e.g., periodic stress timing or mora

2) The difficulty in its most general terms seems to be that a motor execution system that is to interpret specifications in terms of milliseconds would *have to have its own fixed-rate timer* in order to know or specify when *N ms* has expired.

timing). Nevertheless, humans find it very easy to produce speech with a regular periodicity at a global (e.g. phrasal) level, e.g., when chanting, singing or reciting poetry (Cummins and Port, 1998; Tajima and Port, 2003; Port, 2002; Leary, 2003).

Despite these implausible features, one cannot prove the impossibility of such an account. After all, if formal models will implement a Turing machine, they can handle relational temporal phenomena by *some* brute-force method. But an implementational solution along this line is only interesting if specific constraints are applied to the class of acceptable formal models, as Chomsky has frequently pointed out (1965). And, if one can always add additional phonetic symbols with temporal consequences to the universal set and apply as many rules as you please, then proliferation of new universal symbols would undermine credibility.

Yet, short of proliferation of new features, an implementation role for the voice timing effect in English and German cannot be universal. Most languages in the world (including, e.g., French, Spanish, Arabic, Swahili. for example) do not exploit the relative duration of a vowel to the following stop or fricative constriction as correlates of voicing or anything else (Chen, 1970; Port, Al-ani and Maeda, 1980). We know from classroom experience that in cases where English stimuli varying in vowel and/or stop closure duration (with silent, stop closures) lead native English speakers along a continuum from *rabid* to *rapid* -- those stimuli with varying V/C ratio will tend *not* to change voicing category at all for French, Spanish or Chinese listeners. Their voicing judgments are almost completely unaffected by V/C ratio.

They primarily pay attention to glottal pulsing during the constriction. Such durational manipulations may affect the naturalness of the stimuli, but do not make them sound more Voiced or less Voiced for speakers of some languages.

The conclusion we draw from this situation is that English and German manipulate V/C ratio for distinguishing classes of words from each other. English listeners, for example, make a categorical choice between two values of a feature that might be described as 'Voicing' (or as 'Tensity' or 'Fortis/Lenis'). But there is nothing universal about this property. It just happens to be a way that several closely related languages control speech production and speech perception to distinguish vocabulary items. Thus, we have a temporal pattern which apparently must be a learned property of the phonological grammar of specific languages as a 'feature' for contrasting sets of words. To call this distributed temporal pattern a 'symbol,' is to make it impossible to see what it really is - an intrinsically temporal pattern that acts in some contexts like a discrete feature (viz., *Ruby-rupee, bend-bent,* etc.) but which, in other ways, is not symbol-like. For example, it is not static.

To return to the main argument of this paper, such a language-specific, inherently temperal specification for features or phonemes should not be possible according to the formal theory of language. All cross-language differences should be static and segment-sized. And any effects that demand temporal description should be universal. However, there are further problems for the traditional view of phonology and speech timing.

Nondiscreteness in Phonology

Another kind of counterevidence for she Symbolic Phonology hypothesis would be a convincing demonstration of patterns that are linguistically distinct (that is, reproducible and part of the language) and yet not discretely different - not different enough that they can be reliably differentiated. This may seem a difficult sets of criteria to fulfill, but in fact such situations have been demonstrated repeatedly in several languages.

The best studied case is the *incomplete neutralization* of voicing in syllable-final position in Standard German. Syllable-final voiced stops and fricatives, as in *Bund* and *bunt* ('club', 'colorful'), are described by phonologists (Moulton, 1962) and phoneticians (Sievers, 1901) as neutralizing the voicing contrast to the voiceless case. That is, although *Bunde* and *bunte* (with suffixed) contrast in the voicing of the apical stop, the pronunciation of *Bund* and *bunt* seems to be the same, since both words are pronounced [bnt]. The difficulty is that they are not pronounced *exactly* the same (Dinnsen & Garcia-Zamor, 1971; Port, Dalby & O'Dell,1987; Port & Crawford,1989).[3] These pairs of words actually are slightly different as slightly difference as shown in the schematized recorded waveforms in Figure 4. If they were the same, then in a listening task you would expect 50% correct (pure guessing

3) There is at least one other published replication of this effect in Fourakis and Iverson (1984). The magnitude of the non-neutralization effect they observed is very similar to the other studies. However the authors, using certain tests failed to find significance. A sign test across their speakers, however, shows significant differences due to underlying voice.

– like English *too* and *two* would probably show). If different, one would expert at least 99% correct identification under good listening conditions (just like *Bunde* and *bunte* would show). Instead, the two words are different enough that listeners can guess correctly which word was spoken with only about 60–7O% correct performance (Port & Crawford,) 1989). This unexpected level of performance shows that the word pairs are neither the same nor clearly different. The voicing contrast is *almost* neutralized in this context (close enough that both sound "the same"), but toot quite. The differences can be measured on sound spectrograms, but for any measurement or combination of measurements one chooses (vowel duration, stop closure duration, burst intensity, amount of glottal pulsing during the closure. etc.), the two distributions overlap a great deal. If an optimal linear combination of these acoustic measurements is computed (using. e.g., discriminant analysis) then the two classes still overlap so much that it can be classified with 60–7O% accuracy – about the same as native-speaking listeners do (Port and Crawford, 1989)! The Port-Crawford study ruled out the possibility that the difference reflects distorted pronunciations by speakers influenced by the of orthography or that subject were being over cooperative by producing patterns they thought the experimenters wanted to find. This unsettling array of facts led Manaster-Ramer to writ a letter-to-editor in *Journal of Phonetics* (1997) expressing his concern that if the incomplete neutralization phenomenon were correct, then it would imply that linguists could not rely on their own or anyone's auditory transcription. We agree with his concerns: phonetic transcriptions cannot be trusted.

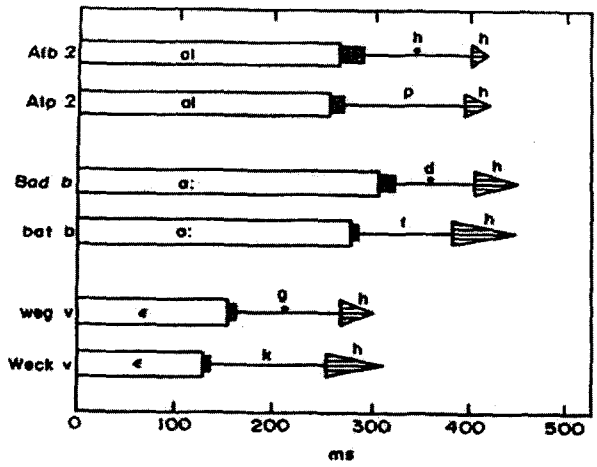

Figure 4. Schematic waveforms of several recorded German word pairs by one of the speakers in Port & O'Dell (1985). The onset of the first vowel (open rectangle) begins at 0 *ms,* the small gray rectangle is the period of visible during stop closure, the straight line is the stop closure which is voiceless, and the triangle represents the stop burst duration (the release of the stop). These results do not support the notion of a static, binary voicing feature [±voice]. While the timing for the voiced and voiceless word pairs are similar, there is a tendency for the vowel before the "underlying" voiceless obstruent (e.g, the vowel in *Alp*) to be shorter then the voiced one. There is more voicing into the stop closure for the voiced stops and also longer visible stop bursts for the underlying voiceless stops than for the corresponding voiced ones.

If this difference is not some sort of artifact, then why phoneticians and linguists have failed to note this before in their transcriptions of German? The answer is that the goal of phonetic transcription is never to record everything, but to record only what is likely to be relevant for a native speaker-hearer (IPA Handbook, 1999). The differences

shown in Figure 4 are too small to be useful in general for communication. (Of course, when taking a forced-choice identification test of minimal pair as Port and Crawford asked their subjects to do, we find that these almost negligible differences can be exploited in perception.) These word pairs lack an essential property of any symbol token (Haugeland, 1981; Manaster-Ramer, 1997; Port, 1997): they are neither discretely different nor are they the same.

A similar phenomenon occurs in American English in the neutralization of pairs like *butting* and *budding* or *writer* and *rider.* The voiceless and voiced stops in *butt, bud, write* and *ride* are, at first listen, neutralized to a flap before the unstressed suffix. But for most American speakers, spectrograms of the two words are somewhat different in that *butting* looks more *t*-like (has slightly longer closure, slightly shorter preceding vowel, slightly stronger burst and less glottal pulsing during the closure) relative to *budding* which is more *d*-like (Fox and Terbeek, 1976; S. Chin,1986). And the percent correct identification in a forced choice task gives a score in the 60-75% range (unpublished data). In both these cases, the English and German speakers are consistently producing a very small difference in articulatory detail. What they produce lies nondiscretely between the two categories (but very close to one value). Other replicated examples of incomplete neutralization are word-final voicing neutralization in Russian (Pye, 1986; Shrager,2003) and Polish (Slowiczek & Dinnsen. 1985). These cases present serious violations of the assumption that phonetics is naturally and universally discrete.[4)]

4) It is interesting and probably important that all the cases of neutralization

One way in the Chomsky-Halle theory to account for the incomplete neutralization phenomena and the essentially temporal cues for features is to postulate a far more finely divided phonetic space -- one that includes differences that can only be reliably identified with 60-70% discrimination. But if the universal phonetic space has that level of detail everywhere, then how could a child's 'transcription' in this alphabet assist in the problem of rapid language acquisition? It would mean that infinitesimal differences in the production of words lead to differences in the transcription. What the child would seem to need is really a much more gross categorization without much detail - so that 20 different productions of the word *cookie,* will all have the same transcription.

Section 3 : An Argument about Linguistic Theory.

The argument we have been presenting can be summarized as follows:

1) The claim that the phonology of languages is a formal system requires, among other things. that there be an apriori **inventory** of phonetic atoms that are **discretely different from each other and** organized into **static segments**, as correctly insisted by Chomsky and Halle, 1968.
2) However, some languages, like English and German, employ **patterns of relative duration** to distinguish classes of lexical items. These contrasts violate the requirement that all distinctive phonetic elements be definable in static terms.
3) Further, some languages, like English, German, Russian and Polish,exhibit

---

mentioned are in contexts where a set of lexical contrast are neutralized. In certain contexts, a distinction is largely lost. It's just that whatever 'process' achieves the neutralization does not completely wipe out the underlying spelling of the lexical items. This is just the kind of problem that a psycholinguistics should be studying.

phonetic categories that are **different from each other but not discretely so**. Even in the same contexts, they largely overlap in the distributions of any measurable phonetic dimensions and are always errorfully identified when hearing any specific token. But they clearly reflect distinct motor patterns. These violate the requirement that all phonetic elements be clearly either the same or distinct.

4) These examples are sufficient to demonstrate that the **phonology of human languages cannot be a formal system**. There are properties of the sound systems of some languages that cannot be described if we are limited to descriptions that comply with the properties of formal systems. Of course, if we conclude that the phonology is *not* formal, there is serious question about whether the rest of language can be strictly formal either.

If a rough approximation to a formal system is all that is required (e.g., for a useful orthography) then a simplified formal model for the phonology of a language may be suitable. But for research into the real nature of phonology, a formal approximation will not do. Its assumptions are too constraining. Only a continuous-time model can take responsibility for both the properties of phonology that appear to be formal as well as for the properties that are not formal.

Counter-arguments and Rejoinders. These two problems may not immediately strike linguists as posing showstopper arguments against the entrenched view that language is a formal system, but we think that, when seriously considered, they present serious problems. One obvious linguistic response to these phenomena might be:

"Your data are only about surface facts, but the formal elements that linguistics studies lie deeper than this. They do not need to be audible on the phonetic surface. Thus, e.g., the incomplete neurtalization phenomena may show that neutralization does not occur in some places where we thought it

did, but the correction for this is simple to postulate a new underlying discrete distinction that happens to be neutralized incompletely during the performance phase at the phonetic output. So there is no deep theoretical problems here."

In rejoinder, we point out two things: this move to pull the language upstairs and out of sight relieves the Symbolic Phonology hypothesis of most testable empirical claims. What had once appeared to be an empirical hypothesis, justified by the phenomena listed in Section 2 above, is no longer subject to empirical refutation. It is true whether or not there are minimal word sets. But this way, the claim that language is formal threatens to become something more like a religious commitment: any incompatible data are dismissed as irrelevant and as revealing a lack of understanding of the nature of language. But this is not a scientifically respectable response. Something more substantive will need to be found to dispute the argument we have presented.

The second problem is that the data we presented above are by no means the only data we might have presented. There is much more.

1) For example, consider voice-onset time (VOT). The famous Licker and Abramson paper (1964) is usually interpreted as showing that there are 3 target values of VOT (in their cross-language summary figure) but the data do not show this at all. It is true that when frequency histograms of measured VOT for word-initial stops are summed across languages, a tri-modal distribution results. But the apparent target durations for the languages do not always lie at these modes. For example, their mean VOT for word-initial aspirated /k/ for a single English speaker is 43 *ms* while the aspirated /k/ in Korean is 125 *ms*. This difference is easily noticed when listening to Korean-accented English. Furthermore, within a

language, VOT exhibits many apparent target values for "aspirated stops" depending on stress, position in a word, place of articulation, etc. (e.g., Lliiker and Abramson,1967; Port and Rotunno,1979; Zue and Laferriere,1979). So there may be certain ranges of VOT that tend to be avoided (if we look only at a single context), but there are still a great many different target VOT values aside from the 3 mythic types: prevoiced, unaspirated and aspirated. Thus far, aside from claim of 4 universal categories of VOT by Chomsky and Halle (1968), there has been no work by phoneticians endorsing *any* specific number of target VOT values. If research in this area has shown anything it is that either speakers actually have continuous control of VOT, or else they employ discrete control using a very large number of categories both within a language and between the languages of the world.

2) Similarly for vowels, Chomsky and Halle suggested 4-5 binary features for coding vowel types (although they proposed that the binary features in principle represent scales in which additional categories are possible). This is probably sufficient when looking only at a single language. But Labov has shown that many historical sounds changes in vowel pronunciation take place gradually by a seemingly smooth shift of target location within a community of speakers (Labov, 1966). The vowel targets of various languages and dialects appear to fall just about anywhere in the F1xF2 plane. Although Ladefoged and Maddieson (1996, pp.4-6) hope it will be possible to specify some universal set of continuous parameters for vowel description, they do not suggest there are only a fixed set of possible vowels. Disner (1983) showed that speakers of two languages with 7-vowel systems located their vowels in slightly different locations in the space. Similarly, the IPA Handbook (1999) speaks of the continuous nature of the vowel space and offers the cardinal vowel system to provide reference points for locating other vowels in this continuous space. In fact, no one who studies phonetics has *ever* suggested that there is a fixed set of vowels across human languages. Despite this, phonologists continue to behave as though there is a fixed universal set of possible vowel types. There are many other examples as well. Certainly intonation shows no sign

> yet that there might be a discrete set of values on any phonetic dimension - whether static tones or contours. Ladefeged has noted that even looking at a feature like implosiveness for stops, there is a gradient between languages in the degree of negative oral air pressure in their production (Ladefoged, 1968, p.6).

The generalization here is that anywhere that you look closely at phonetic phenomena, the cross-language identities evaporate. The /b/ phoneme in English, German, Spanish, etc. are all quite different phonetic objects even if they have many similarities. One might conclude that the only way to believe in a discrete universal phonetic inventory is to avoid looking too closely at the phenomena!

The traditional phonologist might respond:

> *"But none of these observations disprove that discrete features underlie these phenomena. Maybe there are more discrete vowels than we realized. Perhaps we need dozens of VOT values, oral air pressure, intonation coutour, etc., rather the few mentioned in the Chomsky and Halle feature set. So what?"*

We agree, of course, that we have not disproven discrete. apriori features. And if one's only constraint is that there be a finite number, then we will have difficulty winning the argument. But if one postulates a very large and growing phonetic inventory, then, first, one risks the accusation of being adhoc since the theory is indefinitely expandable. But worse is that the use of the phonetic alphabet to account for children's rapid acquisition of language (as envisioned by Chomsky and Halle) strains credulity since repetitions of a single word will become unrecognizable if they have slight differences in, say,

vowel quality, VOT, oral air pressure, intonation, etc. The phonetic alphabet really must be *very small* (that is, let us say, well under a hundred or so segmental features)[5] for it to be plausible a bootstrap for language acquisition.

Altogether then, there is considerable evidence that several essential and unavoidable predictions of the Symbolic Phonology Hypothesis have clear counterexamples. Back in the 1960s, it might have been reasonable to hope that phonetics research would exhibit convergence toward a single universal inventory of phonetic features. But it is clear that 40 years of phonetics research provides ***absolutely no evidence of convergence on a small universal inventory of phonetic segment types.*** Quite the opposite: the more research is done, the more phonetic differences are revealed between languages. So the SP hypothesis and its corollaries were actually disproven long ago and should have been some parts of the phonolput to rest. It must be abandoned as a premise for phonology. Apparentlyogy of individual languages exhibit symbol-like discreteness, but that is as far as discreteness goes. Linguistics cannot make the convenient assumptions of timelessness and digitality for all linguistic units.

The Consequences of the Loss of Apriori Phonetics.

If one were to be persuaded of our primary conclusion so far, that

5) *In the Sound Pattern of English*, Chapter 7, Chomsky and Halle proposed fewer than fifty features. These features have provided the apriori technical vocabulary for generations of research in generative phonology and, more recently, in optimality theory.

'*There is no discrete universal phonetic inventory*,' then what would be the consequences for phonological research? Does it matter? It seems that much of current research in phonology would appear misdirected.

1. **Traditional Generative Phonology** (i.e., the style inspired by SPE) sees its mission as discovering the universal properties of the phonologies of languages. But the search for phonological universals is made much more difficult if there is no universal phonetic inventory. The researcher must now be very careful about drawing any cross-linguistic generalizations. We can no longer assume that the "Voicing" feature in English is the same as the socalled "Voicing" feature in Spanish or Japanese or Arabic. They exhibit some similarities, of course, but still manifest many phonetic differences (Port, Al-Ani and Maeda, 1980). Some of these may be obvious (e.g., the fricativization of voiced stops in Spanish or the aspiration of syllable-initial voiceless stops in English) but others may be subtle requiring experimental methods to see them (e.g., the V/C ratio invariant in English voiced stops). There are still some generalizations to be drawn across languages, such as, say, the tendency of [ki] to evolve historically into [či]. But we cannot assume that any general description, like [Stop]→[Affricate] will have any universal meaning.

Another problem surrounds the time needed to execute rules. Chomsky and Halle dismissed timing measures as irrelevant data. Their theory was about formal relationships, they said. The theory made no claims about timing. The rules they employed are described as 'generative' but are not really executed in time. But the approach endorsed here would disallow that escape. Any theory *must* run in time. Furthermore, the places where discreteness is found now require

an explanation.

2. **Optimality Theory.** Although Optimality Theory (OT) (Pinker and Prince, 1990; McCarthy, 2002) tossed aside much of traditional generative phonology, the new approach remains committed to the principle that language is completely formal and to the mission of discovering the list of linguistic aprioris. For example, OT postulates many operations, for example, *'Gen'*, a component which generates an infinite sets of possible forms given some input, and *'Eval'*, a function which evaluates this infinite set. But these are not to be thought of as operations that take place in realtime anywhere (McCarthy, 2002). Though not part of the stated OT theory, informal statements by OT practitioners reveal that something resembling these formal operations *are* thought to take place over a long period of time during language acquisition and use. At the moment of speaking, however, no rules are applied. The speaker merely selects for production the correct form from a large list of all highest-ranking forms, since all forms have been precompiled, as it were. But the OT approach, like traditional generative phonology, is not defensible if it has operations that do not take place in time.

Furthermore, the entire phonetic feature system from SPE has been adopted without comment into the new theory. All the constraints involved in the ranking process (which make up the empirical content of the theory) are defined using the universal SPE phonetic alphabet. So OT does not in any way escape the concerns raised in this essay about the unconstrained size of the universal phonetic space.

New Directions for Phonology.

We propose that the first steps should be taken toward a *new discipline of linguistics.* Step one must be to naturalize language and fit into a human body, that is, first of all, to cast it into the realm of ***space*** and ***time.*** To do this we must change our focus of attention from the study of *linguistic knowledge* (normally conceptualized as static and symbolic) toward the *study of linguistic behavior and performance.* We do not take this step because of any assumptions about learning or because we deny abstract linguistic knowledge. We study behavior simply because speech and language take place in time. So temporal information is needed to discover how the whole system really works. **Static knowledge of language cannot be separated from the dynamic performance of language**. If the cognitive system for language is something 'designed' to run in time, then it will only be understood in such terms. Chomsky's attempt to separate the static part of language from the dynamic part turns out to do irremediable violence to the entire system. Quite simply language will never be understood by insisting on the distinction between Competence and Performance.

What is *universal* is not any list of sound types, but rather *the strong tendency of human language learners to discover (or create) sound classes.* Humans seek sound types in the speech around them that can be combined in feature-like fashion to specify words. This is what yields all those tables of minimally contrastive words mentioned earlier. The sound system of each language *does* exhibit some discrete features but there is also much that is not discrete or static. Although

we there are some relevant studies of human speech perception that may lead to a plausible psychology on which to base a theory of phonology, a review of this work is beyond the scope of this paper.

It seems likely also that different languages may employ quite different control schemes for speech production. A speech production system must control the muscle systems of the speech apparatus in real time. Although humans all have approximately the same anatomy, the control systems differ greatly (and incommensurably) from language to language. This system is learned by listening, babbling and talking in a language. Again, considerable progress in understanding speech motor control has been made but lies outside the scope of this essay.

Section 4 : Conclusions

We began this discussion of the problem of timing and temporal patterns in human speech by first exploring the theoretical constraints regarding timing that stem from the nearly universal assumption within linguistics that **language is, in fact, a formal symbolic system**. This assumption, which seems so obvious to linguists as to scarcely require any justification at all, turns out to have damaging consequences for understanding how timing could play any role in language and how a discrete phonology could arise from a continuous, noncategorized phonetics.

The evidence against this assumption comes from (a) studies of speech timing showing that some phonologically significant patterns are reasonably described only as essentially temporal ones, So-called

temporal implementation rules cannot provide a reasonable account for them. These violate the premise that phonetic symbols (like all symbols) must be static. The second form of evidence is that (b) some phonological features are not even discretely different from each other. When one hears one of these tokens (e.g., *budding*), it is quite impossible to know with confidence which value of the feature one is hearing. And when you produce one, you cannot tell whether you did it 'correctly' or not. The difference is nondiscrete. Such a situation is another violation of the premise that language is a formal system. In addition, there are many other examples that could also be developed to make this point. These facts imply that, at the phonetic level at least, there are not always apriori, discrete phonetic atoms. If the phonetic space is small, you cannot account for speaker control of speech production or hearers perceptual skills. But if it is large, then you cannot account for language learning.

Symbolic Phonology is based on a metaphor that linguistic structures are made by assembling smaller structure into larger ones. If you are building a horse, it seems you need to start by buying some bricks. The universal phonetic alphabet provides the bricks. This idea once seemed reasonable and perhaps inevitable. But there are other ways to think about the problem of sound structures. There are many ways to construct stable systems from continuous and dynamical components (Port and van Gelder, 1995; Thelen and Smith, 1994; Clark, 1997). Today the assumption that language is completely formal (a) prevents timing from being visible as a property of human languages, thereby rendering irrelevant the research results on the many temporal constraints on phonetic and phonological behavior. (b) forces the

highly implausible assumption that all speech sounds come from an apriori universal segmental inventory. (c) prevents exploitation of data on temporal phenomena (such as processing time, reaction time, response latency, etc.) thereby delegitimizing research in psycholinguistics. Further, (d) it depends on the postulation of a sharp boundary between the formal, symbolic, discrete time domain of language and human cognition ('competence') in contrast to the continuous, fuzzy, realtime domain of human physiology ('performance'). This gap has thus far proven unbridgeable and will remain so as long as the assumption that language is nothing but a formal symbolic system holds sway.

Bibliography

Bannert, Robert (1975) Temporal organization and perception of vowel-consonant sequences in Central Bavarian. *Working Papers 12*, 47-59. (Department of Linguistics, Lund University)

Bloomfield, Leonard. (1926). A set of postulates for the science of language. *Language, 2,* 153-164.

Bloomfield, Leonard (1933). *Language.* (Houghton-Mifflin).

Chen, Matthew (1970) Vowel length variation as a function of the voicing of the consonant environment. *Phonetica 22*, 129-159).

Chin, Steven (1986) Flaps in American English. Unpublished manuscript.

Chomsky, Noam (1959) A review of B. F. Skinner's *'Verbal Behavior'. Language 35,* 26-58.

Chomsky, Noam. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

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

Clark, Andy (1997) *Being There: Putting Body. Brain and World Together Again.* (Cambridge: MITP).

Class, Andre. (1939) *The Rhythm of English Prose.* Oxford: Basil Blackwell.

Cummins, Fred & Robert F. Port. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26.* 145-171.

Dinnsen, D. A. & M. Garcia-Zamor (1971) Three degrees of vowel length in German. *Papers in Linguistics 4,* 111-126.

Disner, Sandra (1983) *Vowel Quality: The Relation between Universal and Language-specific Factors* (UCLA Working Papers in Phonetics 58) PhD Thesis, University of California, Los Angeles.

Elert, Claes-Christian (1964) *Phonological Studies of Quantity in Swedish.* (Almqvist & Wiksell; Stockholm).

Fodor, J. A. (1975). *The Language of Thought.* New York : T. Y. Crowell.

Fodor. J. A. and Z. W. Pylyshyn (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition 28,* 3-71.

Fourakis, M. & G. Iverson (1984) On the 'incomplete neutralization' of German final obstruents. *Phonetica 41,* 140-149.

Fowler, Carol A., Rubin, P., Remez. R., & Turvey, M. (1981). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production* (pp.373-420). New York: Academic Press.

Fox, R. and D. Terbeek (1977) Dental flaps, vowel duration and rule ordering in American English. *J. Phonetics 5,* 27-34.

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision, 12,* 163-186.

Grossberg, S. and Myers, C.W. (2000). The resonant dynamics of speech perception: Inter-word integration and duration-dependent backward effects. *Psychological Review 107,* 735-76.

Halle, Morris, & Stevens, Kenneth N. (1971). A note on laryngeal features. *Quarterly Progress Report 101.* Research Lab in Electronics, MIT 198-213.

Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. (1999) (Cambridge Univ. Press: Cambridge)

Harris, Zellig (1942) Morphem alternants in linguistic analysis, *Language 18,* 169-180.

Haugeland, John. (1985-1981?). *Artificial Intelligence: The Very Idea.* Cambridge. MA: Bradford Books. MIT Press.

Hockett, Charles (1947) Problems of morphemic analysis. *Language 23*, 321-343.

Hockett, C. (1954). Two models of grammatical description. *Word 10*, 210-231.

IPA Handbook (1999) International Phonetic Association, London.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates.*, Cambridge, MA: MIT Press.

Klatt, D. (1976). Linguistic use of segmental duration in English: Acoustic and perpetual evidence. *Journal of the Acoustical Society of America. 59,* 1208-21.

Klatt. D. (1977) Review of the ARPA speech understanding project. *J. Acous Soc. Amer 62,* 1345-1366.

Labov, William 1966. *The Social Stratification of English in New York City.* Washington, D.C : Center For Applied Linguistics.

Ladefoged, Peter (1968) *A Phonetic Study of West African Languages: An Auditory-Instrumental Survey, 2d Edition.* (Cambridge: London)

Ladefoged, P. and I. Maddieson (1996) *The Sounds of the Worlds' Languages.* (Blackwell; Oxford, UK).

Logan, Lively and D. B. Pisoni

Peterson, G. E., and Lehiste, I. (1960). Duration of Syllabic Nuclei in English. *J. Acoust. Soc. Am. 32*, 693-703.

Liberman, A. M. (1967) Perception of the speech code.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press

Lisker, L. & A. Abramson (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*, 384-422.

Lisker. L. & A. Abramson (1967) Some effects of context on voice-onset time in English stops. *Language & Speech 10*, 1-28.

Lisker. Leigh, and Authur Abramson. (1971) Distinctive features and laryngeal control. *Language, 44*: 767-785.

Lisker, Leigh. (1985) Rabid vs. rapid: a catalogue of cues. *Haskins Laboratories Status Report on Speech Research.*

Logan, John. Scott Lively and David Pisoni and (1990) Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Manaster Ramer, Alexis (1996) A teller from an incompletely neutral phonologist. *J. Phonetics. 24,* 477-489.

McCarthy, John J. (2002) *A Thematic Guide to Optimality Theory.* (Cambridge U. P.: London)

Moulton, William (1962) *The Sounds of English and German.* (Univ of Chioago; Chicago).

Newell, A., & H. Simon. (1976) Computer science as empirical enquiry: Symbols and search. *Communications of the ACM 19,* 113-126.

Otake. T., G. Hatano, A. Cutler & J. Mechler. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language,* 32: 358-378, 1993.

Peterson, Gordon E., and Ilse Lehiste. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32:693-703, 1960.

Pind, J. (1995) Speaking rate, VOT and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics 57,* 291-304.

Prince, Alan & Paul Smolensky (1993): *Optimality Theory: Constraint Interaction in Generative Grammar.* Rutgers University Center for Cognitive Science Technical Report 2.

Port, Robert (1979) The influence of tempo on stop closure duration as a cue to voicing and place. *J. Phonetics 7,* 45-56.

Port, Robert F. (1981) Linguistic timing factors in combination. *Journal of the Acoustical Society of America 69,* 267-274.

Port, Robert (1997) The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer. *J. Phonetics 24,* 491-511.

Port, Robert (2002) Phonetics and motor activity. In Fabrice Cavoto (ed.) *The Complete Linguist: A Collection of Papers in Honor of Alexis Manaster-Ramer* (Lincom Europa: Munich) pp.329-344.

Port, Robert F., Salman Al-Ani, and Shosaku Maeda (1980) Temporal compensation and universal phonetics. *Phonetica 37,* 235-252.

Port, R. and T. van Gelder (1995) *Mind as Mition: Explorations in the Dynamics of Cognition.* MIT Press: Cambridge MA.

Port, Robert F., Fred Cummins, and J. Devin McAuley. (1995) Naive time, temporal patterns and human audition. In Robert F. Port and Timothy

van Gelder, editors, *Mind as Motion: Explorations in the Dynamics of Cognition.* (MIT Press. Cambridge, MA), pp.339-372.

Port, Robert and Penny Crawford (1989) Pragmatic effects on neutralization rules, *J. Phonetics 16.* 257-282.

Port, Robert, and Jonathan Dalby. (1982) C/V ratio as a cue for voicing in English. *Journal of the Acoustical Society of America 69*, 262-74.

Robert F. Port, Jonathan Dalby, and Michael O'Dell. (1987) Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America 81,* 1574-1585.

Port, Robert F., and Fares Mousa Mitleb (1983) Segmental features and implementation of English by Arabic speakers. *Journal of Phonetics 11,* 219-229.

Port, Robert and Rosemarie Rotunno (1979) Relations between voice-onset time and vowel duration. *Journal of the Acoustical Society of America 66,* 654-662.

de Saussure, Ferdinand.(1916). *Cours de linguistique générale.* C. Bally & A. Sechahaye, Paris.

Scheutz, Matthias (1999) When physical systems realize functions ··· *Minds and Machines 9*, 161-196.

Shrager, Miriam (2002) Neutralization of word-final voicing in Russian. *Journal of the Acoustical Society of America 112,* No.5, Pt.2.

Sievers, E. *Grundzüge der Phonetik zur Einführung in der Studium der Lautlehre der indogermanischen Sprachen, 5 verb. Aufl.* (Leipzig ; Breitkopf & Hartel).

Sigurd, Bengt (1965) *Phonotactic Structures in Swedish* (Uniskol, Lund). Sweden.

Slowiaczek, L. & D. Dinnsen (1985) On the neutralizing status of Polish word-final devoicing. *J. Phonetics 13*, 325-341

Strange, Winifred (1995) Cross-language studies of speech perception A historical review. In W. Strange (ed.) *Speech Perception and Linguistics Experience : Issues in Cross-Language Research.* York Press, Baltimore.

Tajima, Keiichi and Rebert Port. (1999) Speech rhythm in English and Japanese. In John Local, et al. (ed.) *Papers in Laboratory Phonology VI.* Cambridge University Press. Cambridge.

Thelen. E., & Smith, L. B. (1994) *A dynamic systems approach to the development*

of cognition and action. Cambridge. MA: Bradford Books/MIT Press.

van Gelder, Tim and R. Port (1995) It's about time: An overview of the dynamical approach to cognition. In R. Port and T. van Gelder (eds) *Mind as Motion: Explorations in the dynamics of cognition.* MIT: Cambridge. pp.1-44.

Van Santen, J.P.H. (1996). Segmental duration and speech timing. In Yoshinori Sagisaka, Nick Campbell & Norio Higuchi (Editors). *Computing prosody: Computational models for processing spontaneous speech.* (Springer-Verlag, New York), pp.225-249.

Zue, Victor and Martha Laferriere (1979) Acoustic study of medial /t, d/ in American English. *J. Acous. Soc. Amer. 66,* 1039-1050.

[Abstract]

This paper takes issue with the idea of language as a 'serial-time structure' as opposed to the 'real-time event' of speech, an idea entrenched in Chomskyan model of linguistic theory. The discussion centers around the leitmotif question: Is language constructed entirely from a finite set of apriori discrete symbol types, as the 'competence vs performance' dichotomy implies? A set of linguistic patterns examined in this study, largely with regard to phonological considerations, points to the evidence to the contrary. That is, while the patterns may be said to be linguistically distinct, they are not discretely different, i.e. not different enough to be reliably differentiated. It is demonstrated that much of current research in phonology, including the most recent Optimality Theory, is misdirected in that it falsely presupposes a discrete universal phonetic inventory. The main thrust of the present study is that there is no sharp boundary between 'competence' defined as the formal, symbolic, discrete time domain of language and human cognition on the one hand and 'performance' as the continuous, fuzzy, real-time domain of human physiology on the other.