Modern Generative Phonology: All Form and No Substance?

Peter M. Skaer

1. Overview.

In this paper I explore the issues facing researchers in the scientific study of the sound systems of the languages of the world. This study has been approached from many branches of science, including cognitive psychology, acoustic and phonetic science, generative linguistics (phonology), neuro- and psycholinguistics, and a host of other major and minor scientific disciplines. Each discipline has of course taken its own tact in how and what to study, and most have steered fairly clear of the others in their investigations.

However, over the course of history, these disciplines have run into each other, both intentionally and unintentionally, both literally and metaphorically. In particular, the study of phonetics and phonology have clearly overlapped with each other at different points in their investigations, and have had more than a little impact on each other in terms of understanding the various aspects of the spoken form of human language. Nevertheless, both disciplines tend to work independently of one another, in part due to tradition established by outdated paradigms, and avoid any such interaction as much as possible.

I suggest below that this is not the most productive route to take, and that some collaboration is warranted at certain junctures, and propose that in fact such interaction is not only useful, but necessary. I begin by providing a brief history of the problem, and quickly move to

the present day where I take a look at a few issues from recent research that cannot be characterized effectively solely in one discipline or the other, or where the objectives of the disciplines themselves must be reorganized and/or rethought.

2. Introduction. Form and Substance.

Over the last century or so, the study of the sound systems of spoken language has emerged from rather obscure darkness to a point where much light has been shed on all aspects of human articulatory behavior. During this period of development, a significant body of work has been produced by so-called "generative" phonologists who have sought to develop a minimal set of primitives, and a restricted set of computational rules which act on these primitives, to produce, or generate, an infinite number of surface representations, or examples of spoken language. Generative phonologists have attempted to organize these primitives and rules into *models* of speech production (and sometimes speech perception).

While the study of (generative) linguistics has seen many approaches recently, most will agree that at its foundation, Saussure's (1916) notions of both form (*la langue*) and substance (*la parole*), offer key distinctions which are recognized and adhered to even to this day. Specifically, Saussure stipulated that the "form" (the abstract structure of an utterance) was more important (in the study of language) than the "substance" (the behavioral expression) of an utterance (cf. Lindblom, 2001). We can simplify this somewhat to say that the actual phonetic aspect of a spoken utterance was/is not as important as the abstract model of it. Or, to use Chomsky's terminology, the generative linguist is concerned with the study of, and the development of abstract models of, a speaker's *competence* (what he or she knows), versus his or her *performance* (what he or she does). Chomsky, 1964, states as follows:

> "...the study of actual linguistic performance can be seriously pursued only to the extent that we have a good understanding of

> the generative grammars that are acquired by the learner and put to use by the speaker or hearer. The classical Saussurean assumption of the logical priority of the study of *langue* (and the generative grammars that describe it) seems quite inescapable" (pg. 52).

Modern adherents to this way of thinking are many, but see Hale and Reiss, 2001, for a particularly vociferous support of the position. From a neutral point of view, however, this approach seems questionable at best, since surely the actual physical acoustic parameters of sound production must play a role in the development of any model of speech production.

Ultimately, the goal of phonology "is the construction of a theory in which cross-linguistically common and well-established processes emerge from very simple combinations of the description parameters of the model." (Reiss & Hale, 2002, attributed to McCarthy, 1988). Reiss and Hale clarify this by adding that the goal (of phonology) should be "the development of a theory of possible/impossible human language, not "common (statistically preponderant) human language." (Reiss & Hale, 2002, pg. 2) Obviously theoretical (and abstract) in nature, these definitions do characterize the field of modern generative phonologists adequately. However, a point I wish to raise in this paper is that recently others have taken a more pragmatic approach, such as Massaro, 1998, who states that research in the study of language production and perception "it is necessary to determine how closely the predicted performance matches what is observed and to compare the accuracy of the prediction with other predictions of other theories." (Massaro, 1998, pg. 2) From all of this, my take on things is that we must be mindful of both the practical nature of human speech as well as the more abstract conceptual nature of human speech.

For modern generative phonology, the intention of making a distinction between form and substance in language is to guide us in the determination of what underlies those surface productions, or in other words, determine what the brain does, at least from an abstract

representational point of view, in the production of an utterance. In order to do this, of course, the linguist must observe samples of actual spoken language (the substance), but from these observations, must determine, through abstraction and idealization (cf. Lindblom 2001), representational primitives and computational rules by which the utterance was constructed. The linguist must be able to cut away from variation caused by speakers within a single linguistic group who may speak slightly different from one another due to a host of reasons including variations of oral cavity, pitch, amplitude, velocity, tongue height, lip shape and movement, intensity of vocal cord vibrations, control and understanding of the target "grammar", and so forth. In fact, it is clear that there is much variation in both the production and perception of targeted expressions (be they phonemes, words, phrases or utterances).

As a final note here, before proceeding, I think it is worth at least raising a question on the minimality issue that pervades all of generative linguistics. There is virtual unanimity in the acceptance of the principle that the storage and processing space of the mind is limited, and therefore any good theory of language must restrict the set of primitives, as well as minimize the number of generative rules (or constraints) which interact to produce language. However, to date, there is little evidence that such a mental restriction exists. It is clear that humans have the capacity to store an extraordinary amount of information in the mind, and just using language as an example to illustrate my point, there are many individuals who have demonstrated fluency in not just one or two languages, but in as many as ten or more. They have found space to store numerous grammars, complete with full lexicons, and have not shown any outward semblance of reaching a "saturation" point. As Kirchner points out, "there is no serious psycholinguistic or neurological support for the Generative assumption that mental space is at a premium" (Kirchner, 2001, pg 10). I mention this only because this assumption of limited compacity is ingrained in our understanding of how the brain works,

and how it may be constricted, but the implication of this assumption is directly related to how we build an abstract theory of grammar. We begin with some assumptions that are really untested, though intuitively natural, and such assumptions may in fact lead us astray. While we may, on the other hand, assume that we likely do not store each and every exemplar of language we ever use or hear, it *might* be wrong to think that the mind is strongly guided by the minimalist principle.

3. Variation in Substance.

In this section we will briefly look at a few random but representational issues involving speech production. For speech production, we naturally find variation in how children acquire their native sound systems. In this area of investigation too, we can see that any kind of abstractions regarding the learning behavior algorithms of children require firsthand observations of actual performance data, and that such data may vary greatly within an individual, as well as among a group of individuals. We will then look at a few examples of production variation by adults speaking the same target language, again considering performance data in our evaluations.

Spoken language may vary in obvious ways, such as inter-linguistically, by linguistic code (there are several thousand different languages spoken in the world today), and more significantly for our purposes here, intra-dialectally (there is overtly dialectal variation of course, but even more subtle is the variation that clearly exists between different speakers of a single dialect of a given language). If we are to assume that we can produce a model of spoken language for a given target language, and from a multitude of such studies, produce a universal grammar to which all languages can be reduced, we are thus faced with the task of abstracting the key primitives and computational rules from the data available to us.

We must, however, somehow be able to distinguish beyond

individual idiosyncrasies in speech production. Think for example of how differently a speaker who has consumed 10 pints of beer might say the utterance “Suzie sells seashells by the seashore” as opposed to an equally language-proficient individual who has consumed no alcohol at all. Johnson, Pisoni and Bernacki, 1990, note (as well as many others), that intoxicated speech often contains misarticulations that effect the production of liquids (/r/ and /l/), that often devoice syllable- and word-final phonemes, and that simplify affricates to fricatives, to name just some of the observed changes in performance. This reference to variation found in intoxicated speech is offered simply to define the extreme case of intra-dialectal variation, but clearly, different speakers of the same language do not produce each and every phoneme, syllable and word exactly as every other speaker of the same dialect, drunk or sober. Let us take a look at some illustrations of this point.

3.a. Variation in Child Language.

A very fruitful avenue of pursuit in the study of linguistic variation in speech production is the study of words and utterances produced by children acquiring their native language. We see great variety within an individual, and even more across individuals. For a quick example of this, let us look at several attempts by a single child made at producing the adult target “pen” (/pɛn/), ignoring aspiration (data taken from Faber & Best, 1994).

(1) Variations by a single 15 month old child within 30” period on Adult Target of “pen”

a.	[ma]	f.	[pɪn]
b.	[v]	g.	[thn thn thn]
c.	[dɛdn]	h.	[bah]
d.	[hɪn]	i.	[dhau]
e.	[mbo]	j.	[bua]

It can come as no surprise that a very young child would be found to have such variation in productions of a single target word, even though the child is working within the context of a single target native

language. Yet, do these "performance" -based *substantive* examples tell us anything about how the child's mind works? The traditional generative linguist would have to say no (though many of course have thankfully defied this). In fact, Hale and Reiss, 2001, use the above data to support their claim *against* the use of performance data, on the grounds that it is too haphazard, and could not possibly provide any useful insights or generalizations.

However, from just a cursory review of the data in (1) above, it is clear that the child slightly favors "words" that begin with consonants over those that don't, that final consonants are generally disfavored, that mono-syllabic forms are preferred over multi-syllabic forms[1] (consistent with the target), and that in terms of features, there is an apparent favoring of the labial feature for onset consonants (consistent with the target), and perhaps a slight disfavor of the high feature for vowels. Can these observations be captured appropriately in linguistic theory, or perhaps even more importantly, should they be? I suggest here that indeed, they can and should be, and that with our observations of a single child we can then look at other children, and thus find commonalities in their development, which of course should lend insight into how their minds develop, thereby providing useful insights as to how to best develop an abstract model of human speech and production that allows for development and change, rather than simply standing as a target towards which young speakers aspire. Note, however, that it makes somewhat less sense to speak of children guided by constraints, such as in OT, without at least entertaining some constructive, positive computational rules of grammar. For example, is the child guided by a constraint such as *CODA, ("avoid codas"), or rather, ◎ALIGN(to ONSETS), ("prefer consonant onsets")[2]? I suggest for this tendency of the child's (and I prefer the positive characterization over the negative), and many others, there are both attractive and repellent forces active in the determination of his or her grammar, and it seems therefore somewhat misguided and constrictive to think of a child's productive capabilities only in terms of what

constrains, rather than what drives, him or her. Clearly there is much to learn about how the mind works from the actual speech of an individual child, and from groups of children, within a single homogenous linguistic community, and across diverse linguistic communities.

Let us look at another simple set of children's data. Three different children, all aged eleven months, all born into English speaking families, all attempted to pronounce the word "dog" (see Kawamura, 1998). The results are listed below.

(2) Three different attempts by three different 11 month old children for "dog"

a. [dɔ] Subject T

b. [dɔgi] Subject S

c. [gaga] Subject L

From these three attempts at the English target "dog" (/dag/, or perhaps, /dagi/), we can see possible tendencies. Here, again, we see the preference of CV syllables, with onset Cs preferred over coda Cs, as we saw in our first set of children's data, in (1), but we also see fairly faithful alignment of the qualities of the produced vowel with the first target vowel (all [+low]). However, we also have an additional feature, that of reduplication, either to compensate for the bi-syllabicity of the adult target of "doggie", /dagi/, or simply a manifestation of the child's grammatical propensity at this stage in the children's development. Either alternative is a critical issue in determining what is going in the child's mind. Yet, both end up doing the same thing (copying all or part of a syllable), and thus, there is a danger that without careful attention to the performance nature of this evidence (how, why, when and where it happened), we can be left with simply the output, but no underlying understanding of what motivated such output. Clearly metric mimicry is an issue worthy of consideration, just as clearly as is the propensity to reduplicate in child's, as well as adult's, speech. All a theory of constraints can tell us is "one is preferred over two" (of virtually any phonetic material),

but that "two is really OK" (constraints are violable), and "two is much preferred over three", and so on. It tells us nothing about why the child might be attempting one form or another, since the theory of course does not count perception, or intention, in anyway whatsoever as relevant. Instead, current theory is limited to an idealization of what is said commonly by a representative locus-centered target population. However, by expanding our range of vision to other children, and other cultures, we see that indeed both tendencies (faithful reproduction of syllable onsets of target words, and, partially or wholly reduplicated (doubled) "faithful" reproductions of target words) are present, and must be accounted for within a formal abstract theory of grammar, and grammatical acquisition.

In fact, data from child language acquisition has played a significant role in recent linguistic research, being analyzed by linguists who advocate (consciously or unconsciously) a reduction in the distance between form and substance (see Boersma, 1999, 2000, Hayes 1997, 1999, Kirchner 1997, 1998, 2000, 2001, Lindblom, 2001, Steriade 1999, and others).

Taking one last example, then, let us look at the "toothless paradox", discussed by Boersma, 1999, who quotes Reiss in a private communication: "My reading of the generative linguistics literature is that it is about knowledge states, not behavior. If I don't start flossing, all my teeth may fall out—my pronunciation will change, but my phonology won't" (Boersma, 1999, pg. 11). Boersma addresses this question seriously, and suggests that if we take a child in such a circumstance, one who has lost her two front teeth, for example, we are likely to see a sibilant fricative /s/, formerly pronounced by the child as [s], now pronounced as [ɋ]. See below for a stylized "derivation" of how I imagine this change to take place.

(3) Generative-Structuralist View of Tooth Loss

Before	*After*
/s/	/s/
↓	↓

grammar	grammar
↓	↓
/s/	/s/
↓	↓
body	body
↓	↓
[s]	[ɋ]

Boersma points out that this condition (inability to speak clearly because of tooth loss) is not permanent, nor even long lasting. In fact, well before the child grows her two new adult teeth back in, she has recovered her ability to produce the sibilant fricative, [s], by compensating and altering her articulation in such a manner as to produce a sound that sounds correct, and she does this because of her perception of the target, which has not changed—only her ability to match the target was initially effected, but as Boersma, and others have pointed out, humans are extraordinarily gifted in finding alternative methods to reach heretofore identical goals. There is in fact a great deal of latitude in the articulatory apparatus (cf. Lindblom, 2001), which can compensate significantly in absence of the more direct protocol (speaking with front teeth intact versus minus the front two teeth, for example), or the ability to make perceivable utterances through alternative strategies and obstructions, such as whispering, "Donald Duck" voicing, ventriloquism, speaking with your mouth full, speaking with a cold[3], speaking while biting a pencil, chewing on a pipe, clinching your teeth, and so forth—in fact, the speaker (child and adult alike), has the remarkable ability to compensate in the speech process, which argues significantly towards a goal-guided (perceptually-aligned) theory of speech production, rather than a phoneme-based generative theory of speech production (cf. Weinberg, 1971). The point of this discussion for our purposes here, of course, is the importance of perception in determining targets, and thus, in determining the abstract nature in which language is configured in our brain. For the purposes of this paper we will let these brief examples

suggest what possibly lies behind the use of child language performance data in the development of abstract models of language in the mind. For now, let us turn to an even briefer review of evidence of variation in adult native speakers of a single target language.

3. b. Variation in Adult Language.

First of all, let me acknowledge that there is already significant evidence that adult language varies widely, even under the umbrella of a single mother tongue. We see language forms vary by age, region, gender, amount of education, and so forth. However, the kind of variation I am assuming here starts with a fairly homogenous population, with phonology only as its consideration, and with mutual communicativeness its standard of measurement. So, there is naturally some leeway here, and what is important is that we, as adults, can in fact communicate with virtually all of the different groups mentioned above, as well as any others that might be determined.

Barden et al, 1994, present evidence that speakers do not produce target phonemes exactly the same every time they speak. This makes sense intuitively, yet how does this variation get incorporated into an adequate model of speech production? First let us look at what linguists use to represent sounds in spoken language. At the base, there are of course phonemes. These are symbolic units which essentially represent a combined set of articulatory gestures, referring to lip shape, tongue movement, jaw displacement, vocal cord vibration, and other gestures, depending on how specified the feature matrix is. Importantly, however, these gestures are determined as an all-or-nothing phenomenon, where they are either activated (employed, manipulated...), or they aren't, and are indicated binarily, using a plus or minus feature system, such as [+voice], or [-ATR]. But, can all features really be characterized so easily? What about partial nasals, weak aspiration, devoiced consonants, and so forth? In the binary feature theory, there is no room for middle ground—we can't have, for example, slightly nasalized, moderately nasalized or heavily nasalized

segments (even though phonetically we can see evidence for differing degrees of nasality, as well as for other gradient "features" —cf. Kirchner, 2000). Kirchner suggests that "gradient" rules can be introduced into Optimality Theory (OT), for aspiration (ASPIRATE (gradient)), as follows:

(4) ASPIRATE (gradient): A voiceless stop in initial position in a degree *n* stressed syllable, or in a degree *n* prosodic constituent, is realised with degree *n* aspiration. (Kirchner, 2001, pg. 3.)

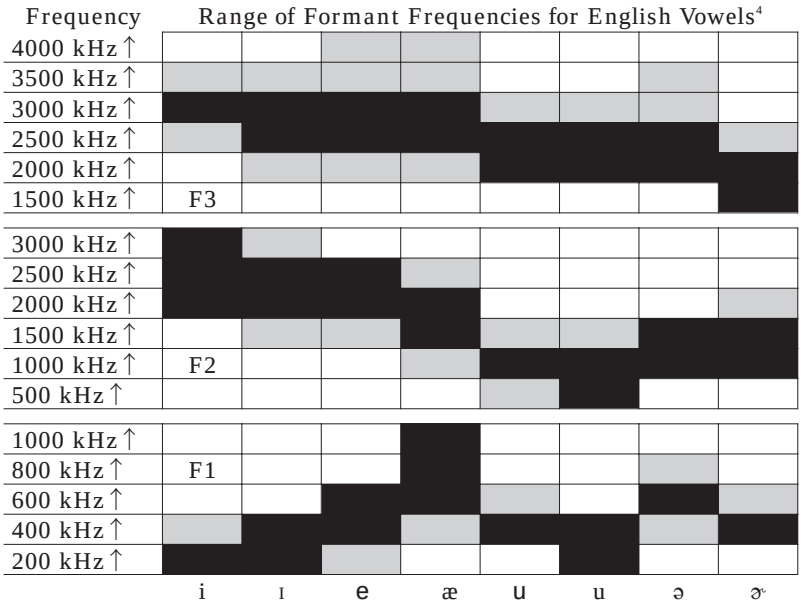
This still does not overcome the faithfulness constraint, PRES (asp), "preserve aspiration", which only responds to the binary features [+asp] or [-asp], and cannot of course accommodate intermediate values of aspiration, as Kirchner himself acknowledges. With standard feature representation, which is restricted within the binary boundaries of current feature theory, full aspiration for word initial stops (such as the /p/ in "peach") cannot be contrasted with partial aspiration for stops placed after an initial-position sibilants in words (such as the /p/ in "speech"). While the idea of "gradient", or "partial", may then also be applied to other features, such as nasality, and consonant devoicing (such as Kirchner suggests for aspiration), it is clear that non-binary feature representations do not fit into even the most current theories very well.

Returning to the physical act of articulation, the results of gestural movement, among other things, is to define the shape of the oral cavity. The shape of this cavity plays an important role in determining the actual sound that is emitted during speech. This is particularly evident in the case of vowel production. The shape of the oral cavity can effect velocity and pressure, and overall, the resonance of the vocal tract. Variation in resonance results in variation in vowel quality, so it is natural that vowels are described in terms of the oral cavity configuration. Vowels that are low and back, ([+low, +back]), for example, have the tongue lowered and drawn towards the back of the mouth, thereby opening up the space in the oral cavity to make it highly resonant, and capable of producing the "deep" vowel sounds of

/a/, or /ɔ/. On the other hand, high front vowels, ([+high, +front]), such as /i/, or /ɪ/, require the tongue raised toward the front roof of the mouth, thereby reducing the size of the oral cavity (since the tongue fills most of it), resulting in the production of thin higher pitched sounds.

All of these movements are *relative* however, since in fact the tongue must be held in some sort of intermediate position, short of touching any extreme (such as the roof of the mouth, or the back of the upper teeth). Because of this relativity, and basic differences in articulatory apparatus, each speaker may in fact produce slightly different variations of the same target. Note the variation and range of the three main formants (resonant frequencies) produced in the articulation of some common English vowels by different English speakers, based on data adapted from Denes & Pinson, 1993.

(5) Variation of Average Formant Frequencies of Some English Vowels



The data in the figure above indicates that there is a wide range of individual differences in the average formant frequencies in the production of representative English vowels amongst the general population of English speakers—yet, most people are able to comprehend each other, and perceive which vowels were being articulated, and which words were being pronounced. In fact there is a significant demonstrated latitude between what is articulated and what is perceived, but unfortunately space limitations prevent us from looking into perceptual issues in more detail here. For now, let us simply understand here that there is a significant degree of variation in specific acoustic parameters regarding what actually constitutes each of the given target vowel phonemes presented above, a fact that any model of speech production must come to terms with.

To this end, nearly half a century ago, Delattre, Liberman and Cooper, 1955, determined that for each phoneme, defined by place of articulation, there was an acoustic locus, or center, around which acceptable phonemes were placed, and the further from the locus the articulated approximation was, the fewer the chances it would be perceived correctly. Others have similarly looked at phonemes as prototypes, or categorial representations, in keeping with the "all-or-nothing" view of phoneme alternation and interaction.

Recently, however, it has become clear to some that distinctions are more relative than absolute, that phonemes simply represent discrete, almost coincidental, target points along a continuum of gestural movement, and that the contrasts themselves (represented by phonetic distance *between* the phonemes) are more important than the actual specific configurations that define or distinguish each specific phoneme. In fact, Barden et al concluded that it was impossible to determine exactly what feature or features, if any, speakers used as the key distinguishing determinators to identify one phoneme from another.

Clearly we do not produce discrete isolated phonemes, but rather a connected stream of phonetic information, just as a leg swings

effortlessly from one point to another in stride. This is significantly different from a freeze-frame approach to gestural activity, which can be likened to movies, where say, 24 frames per second are shown in rapid succession, creating the *illusion* of smooth flowing movement. While for the articulation of human speech, we may characterize the phonetic information by accessing frozen points in time (using phonemes), it is clear that in actuality, the gestures do indeed flow from one to another, seemlessly, with evidence of overlapping articulation and co-articulation involving many components of the articulatory apparatus, which serve to extend the articulation of some property or feature (Flemming, 1995). The question is, do we, as Halle, 1964, suggests, have our analyses completely and utterly depend upon the notion that speech is comprised wholly of phonemes composed of distinctive features which represent a sequence of discrete entities, at the exclusion of all other phonetic detail available in an given utterance?

It is clear that phonetically, there is smooth and continuous movement from one phonemic target to the next, that there is upward and downward movement of jaw and tongue during the production of a string of articulatory segments, as well as gradual, rather than drastic, releases of vocal cord tension, nasal cavity closures, and so forth. There are in fact few leaps from one position to the next, instead, there are usually smooth transitions, with some features coarticulated at certain junctures, while others not. Note Kirchner's, 2002, description of averaged fundamental frequency values of the utterance /la/, sampled every 10 msec., (6.a), compared to the second version, which is not attested to, and yet indicative of a discrete phoneme theory, with no transitions from one phoneme to the next (6.b).

(6) Syllable /la/, sampled every 10 msec[5]. (adapted from Kirchner, 2002, pg 23)

a. 100-106-112-118-124-130-136-142-148-154-160-166-166-172-178-184-190-196-202-208-214

b. 100-100-100-100-100-214-214-214-214-214-214-214-214-214-214-214-214-214-214-214-214

llllllllllllllllllllllllllllllllllllllllaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

In (6a), above, we see the gradual transition from a low frequency to a slightly higher frequency that takes place throughout the duration of the spoken exemplar. Essentially each "snapshot" taken during the utterance will render a slightly different average fundamental frequency. There is a clearly observed graduation from the beginning value to the final value, with noticeable discrete minute differences detectable at virtually every stage of the production. However, present generative phonological theory ignores all of the intermediate stages, and presents us with just two "snapshots"; one of the phoneme /l/ at one average frequency, and the other of the phoneme /a/ at another average frequency. Clearly this idealistic view of production does much to distort our true understanding of what actually happens when this sequence of sounds (or any other) is produced.

To further illustrate this point, Steriade, 1997, provides numerous examples where the timing of one gesture does not exactly coincide with another in discourse, where one begins earlier, and ends earlier, than another, for one example. This would suggest that at any one discrete point in time, a given feature may exist ([+F]), and at a later point in time, before articulation is concluded, that same feature will not exist ([-F]). Let us look at one example for clarification.

(7) Laryngeal Gesture Timing (Steriade, 1997, pg. 16)

a. Peak of laryngeal gesture timed to onset of oral constriction: e.g. ^{h}t

[-----glottal abduction-----]

[-----oral closure----release----]

b. Peak of laryngeal gesture timed to onset of oral constriction: e.g. t^{h}

[-----glottal abduction-----]

[-----oral closure--------release-]

Depending upon where we take the phonemic 'snapshot', both (7.a) and (7.b) could be described as having the features [+g.a., -o.c.], or [-g.a., +o.c.][6], with neither representation offering a good basis of understanding what is really going on in the two clearly distinct vocalizations.

Another thing clear about phonemic descriptions is that there is often an excess amount of information within a complete distinctive feature matrix of any phoneme, much of which is predictive, redundant or simply wrong (Keating, 1984), and yet, at the transition points between phonemes, there are features (or gestures) which may be extended, or overlap, and other vocalization features, such as voice onset times, ascending and descending states, releases, and silences—all of which are phonologically relevant (cf. Saussure 1916: 79–95), but are not considered at all, within most current phonological frameworks, though this information can have a significant impact on perception of target phonemes and/or words.

Evidence of variation from adult performance in *casual* and *fast* speech also is rich with insights that support the use of performance data in developing models of language processing in the brain (see Skaer, 2001, for a consideration of these processes in light of an effort-based theory of phonological production). From this section, it should be clear that there is much variation in human speech, both for children and adults, and that generalizations solely based on the idealistic concept of *form* for human speech at least partially obscure the rich and fruitful data available to us from actual speaking situations, the *substance*. This, further attested to with the notion of a discrete phoneme, which in fact in actual speech is anything but discrete, suggests that there is some room for improvement in the realm of phonological theory.

4. Conclusion.

Clearly speech is part of a dynamic chain of physical events, and just as clearly, it must depend on, at least partially, the overall methods and constraints that guide the rest of the physical body, employing basic principles of conservation of energy through ease of effort, and obeying natural physical laws of energy, propulsion and resistance, while embodying unique artifacts of language evolution that allow for both systematic and non-systematic occurrences of

language forms and other subtle aspects of universal and individual language evolution and variation.

We can think of language as made up of a number of primitive elements, basic units of action, that are then assembled into more complex programs. Surely we would expect some correspondences between these primitive motor elements and the basic primitives of human speech (the phonemes), and further we would expect convergences in how the computational rules may mirror neural "rules" governing the organization of motor units into a "theory" or grammar of complex (linguistic) behavior. (see Cortona, 1988, for a related discussion on the *motor theory* of language).

The separate notions of the form and substance of language, however, has caused many linguists to ignore important linguistic information that is obtainable from the performance side of language, in the development of new and evolving theories of how language processes may be thought to represent the complex inter-workings of the human mind. Currently, many phonologists have adopted Optimality Theory as the best model of language produced so far, at least for describing, explaining and understanding how language is spoken. Notably however, OT may not be the best tool for this since OT is not really a model of grammar, but an ever-increasing list of prohibitions—what you can't do is hardly a prescription for what you can do—and to date there is very little to suggest that proponents of the theory have come anywhere near grappling with this basic fact of OT, that the set of candidates generated by the "grammar" (of which there is actually none), the possible permissible strings before evaluation, is indeed unbounded, or infinite—so unbounded that a grammar to be evaluated could even include non-linguistic information as well. Hale and Reiss correctly point out that "no language marks past tense by having the speaker eat a banana after uttering each verb" (Hale and Reiss, 2001, pg. 13). Well, we all know that such a possibility is embarrassingly implausible[7]. We know that a given theory is intended to fit a particular domain of study, which is in this

case (OT), that of the study of human discourse, and thus, we dutifully restrict our choices to only those possible outputs that could likely qualify as language exemplars. And then, if for no other principled reason than for brevity's sake, we further limit ourselves to strings that at least have a chance of being possible survivors of the evaluation process.

Yet of course this speaks of a grammar we all know to exist, that is in fact not built up of negative constraints, but rather is productive, comprised of at least an elastic sense of underlying primitives and generative, productive computational rules which can be used to develop candidates which may then compete with one another depending upon context (in the case of phonology at least), and type of delivery. Unfortunately, OT does not stand up even to the most rudimentary tests of scientific inquiry, which begins with one or more positively stated generalizations, which can then be tested against other theories, and against itself. We need a model that contains at least some idea of what we have to begin with, and what direction it is heading, before we can start constraining our productive capabilities by establishing and ordering a set of constraints—they first have to apply to some manifest entity (a posited "grammar") to have any inherent plausibility of their own.

We end this discussion here faced with two rather overwhelming conundrums. On the one hand, we currently have a theory (OT) that has no productive metaphor, no positive propensity to produce productive output, thus allowing a literally "anything goes" grammar base, with an ever growing list of violable constraints, which together act as a powerful, nearly magical, filter system which allows one single successful candidate through to the surface. The form of language, as such, lacks a skeleton, a backbone, or a central unifying structure. And yet, on the other hand, we are prevented from observing the real flesh of language, the actual spoken utterances of children and adults alike, and left to imagine and idealize what language is—given these conditions, we can hope for no better than coincidental conclusions,

partial truths, and half solutions. I suggest that both the form and substance of language are two parts of the greater whole, and together make a vibrant, organic dynamic body capable of speech and a great many other things. It seems to me, that in order to understand humans, their brains, and their language, we must unite, rather than divide, our concepts of form and substance, competence and performance, brain and body, phonetics and phonology.

References

Barden, G., K. Harris and L. Raphael. 1994. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. 3rd. Ed. Baltimore, MA: Williams & Wilkins.

Boersma, P. 1999. On the need for a separate perception grammar. Ms. http:/www.fon.hum.uva.nl.paul/

Chomsky, N. 1964. *Current Issues in Linguistics Theory*. The Hague: Mouton.

Cortona. 1988. *The Motor Theory of Language: Origin and Function*. Ms. (see Robin Allot Homepage)

Dang, J. & K. Honda. 1996. Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. *Journal of the Acoustic Society of America*, 100, 3374–3384.

Delattre, P., A. Liberman & F. Cooper. 1955. Acoustic Loci and Transitional Cues for Consonants, *Journal of Acoustic Society of America*, 27: 769–773.

Denes, P. & E. Pinson. 1993. *The Speech Chain*. New York: W.H. Freeman & Co.

Faber, A. & C. Best. 1994. The perceptual infrastructure of early phonological development, in Lima, S. (ed), 1995: *Reality of Linguistic Rules*. John Benjamins.

Flemming. E. 1995. *Auditory Features in Phonology*. Doctoral Dissertation, UCLA.

Hale, M. and C. Reiss. 2001. Phonology as cognition. Concordia University, Montreal. Ms.

Hale, M. and C. Reiss. 1998. Substance abuse and dysfunctionalism: Current trends in phonology, paper presented at the *Montreal Ottawa-Toronto Phonology Workshop*, 1998.

Halle, M. 1964. On the basis of phonology, in Fodor J. and J. Katz (eds), 1964: *The Structure of Language*, New York: Prentice Hall

Hayes. 1997. Phonetically driven phonology: The role of Optimality Theory and inductive grounding, paper presented at the *Milwaukee Conference on Formalism and Functionalism in Linguistics*, 1996.

Johnson, K., D. Pisoni and R. Bernacki. 1990. Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica*. 47: 215–237.

Kawamura, M. 1998. *The Early Phonologies of Japanese and English Speaking Children*, Hiroshima University Senior Thesis.
Keating. P. 1984. Phonetic and phonological representation of stop voicing. *Language* 60.
Kirchner, R. 2002. Contrastiveness and Faithfulness. (to appear in *Phonology*).
Kirchner, R. 2001. Preliminary thoughts on "Phonologisation" within an exemplar-based speech processing system, to appear in *UCLA Working Papers in Linguistics*, vol. 6.
Kirchner, R. 2000. Phonological contrast and articulatory effort. University of Alberta Ms.
Kirchner, R. 1998. *An Effort-Based Approach to Consonant Lenition*. Doctoral Dissertation, UCLA.
Lindblom, 2001. Developmental original of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. Ms. Stockholm University. Sweden.
Massaro, D. 1998. Categorical perception: Important phenomenon or lasting myth?, paper given at *International Congress of Spoken Language Processing*, Sydney, Australia, 11/30–12/6, 1998.
Reiss and Hale, 2002. The comprehension/production dilemma in child language: A Response to Smolensky. Concordia University, Montreal, Ms.
Saussure (de), F. 1916. *Cours de Linguistique Generale.* Paris: Payot.
Skaer, P. 2001. An introduction to the phonology of casual spoken English. *Hiroshima University Journal*, Volume 27: 23–49.
Steriade, D. 1997. Phonetics in phonology: The case of laryngeal neutralization. UCLA Ms.
Vanderbilt, 1988. *Motor Theory of Language Origin and Evolution.* Ms. (see Robin Allot Homepage)
Weinberg, B. 1971. A study of buccal-speech. *Journal of Speech and Hearing Research*, 14, 657–658.

Note

1 An anonymous reader suggests that this observation is "untenable" since the input is monosyllabic. However, as Kawamura, 1998 demonstrates, there is a strong tendency for children to reduplicate mono- and multi-syllabic targets, so I will allow this observation to stand since I am simply suggesting *possible* patterns that may be observed in the substantive data of actual child language productions.

2 I have used the target symbol, "◎", to indicate an attractive force, as opposed to the asterisk, "*", which indicates a repellent force, or a constraint.

3 Dang and Honda, 1996, observe, for example, that "the shape and size of [nasal] cavities vary widely due to the swelling of mucous membranes during upper respiratory tract infections 'colds'. Such variability is important in shaping the speech stream as it introduces variable 'antiresonances' in speech, and changes the spectral shape of nasal formants." (pg. 26–27)

4 In this figure, more than 25 samples of each vowel were analyzed; the gray

cells represent 5 or fewer occurrences in the given frequency range, with the black cells representing 6 or more occurrences. White, or blank, cells of course represent no occurrences. —see also Barden et al, 1994, for similar variation data.

5 Kirchner lists 21 10-msec increments in his example, as shown here, which would result in the length of the utterance at about 210 msec. However, based on my own tests, I find this to be on the short side. I observed 10 sample utterances of /la/ which produced lengths from 206 msec to 549 msec, with an average of 430 msec—thus we would expect even more discrete gradient samples to be added into the fold.

6 Where "g.a." stands for *glottal abduction*, and "o.c." stands for *oral closure.*

7 Or is it? Clearly there is much research that suggests that gestures play an integral role in the completion of a speech act, with gestures sometimes acting as a pantomime that accompanies the spoken utterance, sometimes as words and utterances themselves. Recent studies in motor theory have suggested even more direct, and constant, links between speech and gestures. (see Vanderbilt, 1988)