Functionalism and Formalism in Linguistics

VOLUME I: General papers

Edited by
Michael Darnell,
Edith Moravcsik,
Frederick Newmeyer,
Michael Noonan,
Kathleen Wheatley

Offprint

Formalizing Functionalism

Janet Pierrehumbert
Northwestern University

Abstract

As the companion discussant paper to the position papers on phonology by Bybee and Hayes (this volume), this paper first notes points of divergence and convergence of the aforementioned authors' positions. It is then argued from the viewpoint of experimental phonetics that in two areas, granularity and dimensionality, further theoretical work is required to 'fully formalize' their theoretical positions.

1. Introduction

The organizers of the conference have assigned me, perhaps to their surprise, the job of commenting on papers by two functionalists, Bybee and Hayes. Bybee is known in formal circles as a functionalist (though she disputes this designation) and Hayes has taken up the functionalist cause by pursuing the general program of Ohala and Lindblom within the framework of Optimality Theory. I am happy to comment on two functionalists, because my interest in both formalist and functionalist work on sound structure is based on the extent to which both bring us closer to the Holy Grail of scientific theory — a model which is incisive, comprehensive, fully formalized, and exhaustively validated. With much of my own work being in experimental phonetics, I take "fully formalized" to mean formalized down to the last differential equation, and "exhaustively validated" to

reflect coverage of all speech behavior in its full statistical variability and physical glory.

I use the word "formal" in the simple and original sense it is used in mathematics, including both calculus and logic as examples of "formalism." In contrast to Pullum (1991) or Anderson (this volume), I feel it is counterproductive to restrict the term "formal linguistics" to work in linguistics which is formalized using the resources of logic and formal language theory. The term properly refers to work in linguistics which is formalized by any technically sound means whatsoever. It appears entirely probable to me that the ultimate theory to which I just alluded will owe as much to Pascal, Descartes and Newton as it does to Aristotle, Russell, and Turing. Therefore, I see no intrinsic contradiction between functionalism and formalism, and I will take advantage of the substantial amount of agreement between Bybee and Hayes to raise some issues about the formalization of functionalism.

Let me begin by mentioning some important points on which Bybee and Hayes agree. Both agree that phonological constraints (such as *[+nasal][–voice]) are schematic descriptions of forms. They are descriptions because they describe the sound patterns of words. They are "schematic" because a great deal of detail is missing, compared to the full description of any particular word token. Hence, they bear something of the relationship to actual word productions that a sketch bears to the visual scene it describes. Bybee and Hayes also agree that phonetic properties provide the vocabulary for phonological constraints. Features such as [+nasal] and [–voice] denote complexes of articulatory and acoustic observables. They further agree that phonetic generalizations are a source of constraints on co-occurrence and sequencing. That is, phonetics is reflected not only in the segmental or featural inventories of languages, but also in syntagmatic structure.

A last point of agreement between Bybee and Hayes is that phonetics does not provide the whole story for phonology. Nonphonetic factors, such as symmetry and morphologically driven analogy, may induce regularities of sound structure which are phonetically unnatural.

The conclusions I just summarized are true, and you should believe them. They did not come cheap, however obvious they may now seem; they are the fruits of a tremendous scientific effort carried out over decades. Many of the key papers leading to these conclusions were written by Bybee and Hayes themselves. There is no need to devote further space to discussing them here.

Next to these important points of agreement between Bybee and Hayes, there are also important differences in the viewpoints they have presented. They differ in their conception of "functionalism". Hayes describes languages as

optimizing contrastiveness and ease of articulation. A particularly noteworthy aspect of his view is the claim that the optimization is local rather than global. That is, phonological grammars may in effect come to rest on a local optimum which is not a global optimum. In addition, he adds symmetry to the phonetic functions of ease of articulation and acoustic contrastiveness, which have been so intensively studied by other functionalists such as Ohala and Lindblom. Bybee specifically takes issue with the view of functionalism which holds that language is in any serious sense optimized. Rather, her claim is that the structure of language is shaped by actual instances of language use, and she proposes the term "usage-based phonology" for the approach to phonology that she has in mind. Because of her emphasis on usage, frequency and generalizations over the lexicon have a role in her approach which they lack in the proposal presented by Hayes.

2.   Where Constraints Come From

According to Bybee, constraints are (implicit) generalizations over actual speech tokens. This suggestion may be interpreted in two ways. It is possible that individuals form these generalizations as a result of their experiences in speaking and listening, and represent these generalizations unconsciously in their minds. But, Bybee's analogy of the checkout lines in the grocery store raises the possibility that constraints cannot even be imputed as such to the minds of speakers. Instead, they could represent external scientific generalizations about the patterning that results from speech processing. Even allowing for this more radical possibility, it still seems appropriate to speak of phonological "constraints" in much the same way that one might view the laws of statistical thermodynamics as "constraining" the physical universe. Whatever dynamics in the brain may indirectly result in phonological laws such as the laws of syllable structure, it is still the case that these pervasive laws constrain our models. As scientists, we consider models which exhibit these patterns in their long-term behavior, and we can eliminate those that do not. In this sense, we are still justified in imputing the phonological constraints to the cognitive system.

With constraints held to emerge from the aggregate properties of many experienced tokens of speech, it is obvious that the nature of the constraints will be formed by the nature of the phonetic material. Bybee's present position thus develops directly from her earlier work as one of the founders of Natural Phonology. Speech consists for the most part of words, and children speak in

words long before they have mastered the adult phonology. As a result, Bybee's proposal integrates lexical information with physical constraints right at the ground level. That is, constraints are not founded merely on some kind of universal phonetic experience, but on language-specific phonetic experience accrued while hearing and producing a given language. This provides an important contrast with Hayes' work which I will take up below. Bybee asserts that repeated use of similar patterns results in stereotyping. A more explicit and formalized proposal is needed to fully evaluate the force of this idea.

Hayes has a very different conception of how phonetically grounded constraints arise. According to his theory, candidate constraints are generated top-down by creating all possible combinations of the primitives of phonological theory. These primitives are deemed to be innate, having arisen by evolutionary adaptation. Hayes' view of what these primitives are is a standard one within generative phonology; the primitives of his model include features, structural nodes, and the relations of dominance and temporal precedence. This means that the candidate constraints use arbitrary fragments of hierarchical phonological representations. Given the large set of such constraints, Hayes posits a procedure for weeding out constraints which do not represent a local optimum. This procedure applies iteratively between constraints which are neighbors, where he defines the neighbors of a constraint as those which differ through a simplification or change in one particular. A constraint which differs through addition of a particular does not count as a neighbor. In short, a constraint is eliminated if it has as a neighbor a constraint of equal or less complexity which is more effective in describing the map of phonetic difficulty. The surviving constraints are then ranked using the constraint ranking made available in Optimality Theory.

Hayes devotes a smaller part of his paper to phonetically ungrounded constraints, which may arise in languages through a "conspiracy of historical circumstances", such as a leniting change followed by a massive set of borrowings. His suggestion about these constraints is that children learn them from generalizations over the input data. "Eventually, the child figures out such constraints from (systematic, consistent, long-term) negative evidence" (draft p. 26) As a corollary, he further predicts that only grounded constraints could influence intuitive judgments. This prediction is directly opposite to the prediction of Bybee's model, according to which phonetic and lexical information is homogenized in the projection of constraints.

3. Granularity

In evaluating Hayes' proposal, it is important to bear in mind that the phonetic map he presents as an example is extremely rough and schematic. Real phonetic maps involve continuously variable functions which are described formally using calculus. Furthermore, the phonetic expression of features is known to be language and context dependent. For example, Keating (1984) shows that the feature [+voice] for stops is expressed differently in word onset than in medial (intervocalic position), and differently in English than in Polish. Comparing her results to those for Urdu in Hussain & Nair (1995), we note that in word onset position, English expresses [+voice] for stops mainly in the release and vowel onset, whereas in Urdu, the information is phonetically localized in the closure. Some stops that would count as [+voice] in English would count as [−voice] in Urdu. Obviously, the tradeoff of [+voice] against features governing place of articulation and against the features of the preceding and following vowels will depend on where the expression of [+voice] is localized. In taking [+voice] to be an innate universal feature, Hayes has thus left open the question of what region of any given phonetic map counts as [+voice] and what region counts as [−voice]. He has not told us anything about how the language learner establishes the truth conditions for the feature on the phonetic materials in his or her experience. In the light of this problem, I would say that Hayes' proposal goes less far towards bridging the gradient and continuous character of phonetics to the granular character of phonology than it at first appears to.

Hayes also has a serious problem with granularity right within the phonological description. In his example, the winning constraint (*[+nasal] [−voice]) is at the most specific level considered, and it has perfect effectiveness. In a realistic physical model, it is almost impossible that a constraint as coarse as this would be perfectly effective. Furthermore, this constraint would be in competition with vastly more specific constraints which specified the place of the constriction for the nasal, the other features of the voiceless consonant, the prosodic and morphological position of the two consonants, and so forth. Hayes' comment on this point ("I have imposed a relatively arbitrary limit of formal complexity on this candidate set, under the assumption that language learners either cannot or will not posit extremely complex constraints") appears to me to be unconvincing. Many phonological constraints are far more complex than the ones Hayes presents. Language learners do learn individual lexical items, which according to both Lexical Phonology and other more recent theories (such as Declarative Phonology) are constraints on individual forms. In the domain of syntax, models of human parsing

performance can be considerably improved by imputing to the mind recollections of quite complicated structural and lexical templates (see Bod 1995).

Since fine constraints would sample the phonetic map more finely than coarse constraints they would model it more accurately and would have higher effectiveness values. Thus, a correct fine constraint would survive in addition to more general competitors under the algorithm Hayes proposes. The final list of constraints would therefore contain all correct constraints at the finest level that the algorithm projected in the first place. The algorithm does not in itself propel the phonology towards a description which is more granular than the phonetic one. All of the work in this regard is born by unstated presuppositions about which constraints are considered at all.

In raising these problems with Hayes' proposal, I do not mean to suggest that Bybee's approach offers an immediate solution to the problem of granularity. In fact, the same problem comes up again in a different guise. Modeling the empirical results that Bybee presents obviously depends on the assumption that the cognitive representations of words have an immense amount of phonetic detail. Consider the continuum represented by: full vowel+r... schwa+r ... syllabic r ... nonsyllabic r. Each word such as "mammary" or "opera" would, in Bybee's view, have associated with it a probabilistic distribution over this phonetic continuum, and this distribution would be incrementally updated whenever the word was used. Given the amount of progress which has occurred in statistical pattern recognition techniques, I will view as unproblematic Bybee's assumption that phonetically diverse productions of the same word can be properly assigned to the cluster that they belong to. That is, we will assume that all tokens of the word "opera" can be recognized as such, so that their properties do indeed update the correct probability distributions. What remains unclear is why the whole behavior of the system is more granular as we move from the post-lexical phonology to the lexical phonology.

As an example, consider the data on the strong and weak grades of Sanskrit roots, as presented in Steriade (1988). The strong grade of a root typically has /a/ in nuclear position. In the weak grade, this /a/ is not present and the nuclear position is filled by a consonant which appears in coda (or occasionally, an onset) position in the strong grade. For example, for the root whose strong grade is /smaj/, the corresponding weak grade is /smi/. In /smi/ the nuclear position is occupied by the offglide of the strong grade.

According to Steriade, the availability of the weak grade depends on the sonority of the coda consonant of the stem. The dependence is summarized in Table 1, computed from raw counts in Steriade (1988).

Table 1. *Percentage of Stems Having a Weak Grade Form.*

Coda consonant			
/m/	/n/	Liquid	Vocoid
30	75	88	99

/m/ is less sonorant than /n/ in Sanskrit, according to Steriade. Steriade reports a similar effect of the sonority of onset consonants and asserts that the onset and coda sonority interact cumulatively in determining the probability that a weak grade form exists.

These data present a close analogy to those discussed by Bybee, because they display a regular relation between frequency of occurrence and a phonetic continuum. There is an important difference, however. Individual stems in Sanskrit do not alternate. A root either has a weak grade or it doesn't. The statistical regularity is not over tokens of words, but rather over types of words in the lexical inventory. The regularity is very coarse-grained compared to those discussed by Bybee; the data fall into separate bins and all the in-between cases found in the study of lenition are not found here.

In principle, lengthy iteration of a probabilistic process can create categorical differences from noncategorical ones. An example would be models of speciation in evolutionary biology, in which probabilistic genetic mutations can interact with selectional pressures to create separate species from the variable population of a single species. I believe that something along these lines is what Bybee would suggest for the Sanskrit data. A specific formalized proposal, even of a rather schematic character, would be a most welcome contribution towards clarifying the consequences of her approach.

4. Frequency and the Lexicon

As noted in the first section, Hayes draws a distinction between phonetically grounded constraints, which are projected without reference to the lexicon, and phonetically ungrounded constraints, which are projected as negative generalizations over the lexicon. This separation between the grammar and the lexicon is taken from Optimality Theory, which he advocates as the overall framework for his proposal. Standard Optimality Theory (as laid out in, e.g., Prince and

Smolensky 1993) draws a strong distinction between the grammar and the lexicon. Constraints are universal and grammars of individual languages rank them. The lexicon provides candidates via the Gen function, and Faithfulness (or more recently, correspondence, constraints) provides an interface between the evaluation of alternatives and the lexical inventory itself. This distinction is highly reminscent of the Chomsky and Halle (1968) model. In maintaining it, Optimality Theory stands as an exception to the general trend in linguistic theory towards lexicalist approaches. This trend otherwise encompasses developments as diverse as Lexical Phonology, Declarative Phonology, LFG, and Minimalism.

I cannot agree with the separation of the grammar from the lexicon which Hayes appears to presuppose. Studies of language acquisition show that the child's acquisition of phonology is mediated by the acquisition of actual words. Children extend their vocabularies above all by adding new words which are phonologically similar in various regards to words they already know. In short, they take advantage of articulatory routines that they have practiced in actual words, and by this method, they incrementally progress from phonologically simple forms to more complicated ones. (Edwards 1996; Lindblom 1992; Ferguson and Newport 1975). In adult judgments of acceptability, acceptability is directly related to lexical frequency, contra Hayes' suggestion quoted above. (See Frisch 1996; Treiman *et al.*1996; Pierrehumbert 1994; Coleman and Pierrehumbert 1997).

A specific and very telling problem with separation of the grammar from the lexicon is that it leads to difficulty in characterizing systematically underrepresented but still extant forms. Berkley (1994b) discusses this problem as it manifests itself in OT, in connection with the productivity of affixation. Through detailed study of an on-line English dictionary, she shows that attachment of the suffix "ity" is systematically but not absolutely suppressed on otherwise eligible stems which end in coronal obstruents. That is, although "chastity" and "sanctity" exist as words, many other similar forms such as "distinctity" are not attested. The degree of underrepresentation of such forms is shown to be statistically significant, and this provides evidence for the activity of OCP-Place in the phonology. Now Optimality Theory can, like any other theory, describe accidental gaps in the word inventory in terms of accidental gaps in the set of available lexical representations. It has two resources for describing systematic gaps such as *distinctity. The first, presented in Prince and Smolensky (1993), is the use of a constraint ranking which causes the null parse to emerge from EVAL as the preferred outcome. A null parse means that the word can never be pronounced. A second line of explanation would hold that the best

outcome for "distinctity" is identical to the more faithful outcome of some other underlying representation. In this case, the form would not be learnable as a distinct lexical item. For example, if (by virtue of some remarkable constraint ranking) the best phonological outcome for "distinctity" were "distinctness", children acquiring English would have no reason to learn "distinctity" as a separate lexical item from "distinctness".

The problem noted by Berkley is that any sensible constraint ranking which achieves such an outcome for "distinctity" would also achieve an analogous result for "chastity". To generate "chastity" without also generating "distinctity", it is necessary to associate a separately ranked grammar fragment with each of the words in question. Berkley views such a proposal as a reductio ad absurdum of the line of explanation. However, Hammond (this volume) does in fact advocate handling exceptional lexical items by inserting sets of item-specific constraints into the constraint hierarchy. I view Hammond's solution as undesirable because it provides no insight into the statistical relationship between the low rate of occurrence of such forms and the phonetic properties which gave rise to this low rate. It at best mechanically encodes observations such as those in Table 1 without treating them as a deep and characteristic regularity of language. As a general solution to statistical regularities in the lexicon, it also runs afoul of the fact that OT constraint hierarchy is only one-dimensional; see the following section.

Bybee's recent work (see also Bybee, forthcoming) develops this issue much further by demonstrating the existence of regular relationships between continuous phonetic parameters and the probability distributions of word pronunciations. Neither the phonetic parameters nor the probabilities are viewed as logically prior; the observation is only that a regular relationship obtains between them. The data presented in Bybee's papers concern the statistical distribution of allophonic details as a function of the frequency of words in running speech. For example, the mode of the statistically variable process of /t/ reduction is shown to be monotonically related to the frequency of the word in which the /t/ occurs. This example provides, according to Bybee, a case study of a general property of language: the extent and frequency of leniting allophony are a direct function of word frequency. As I have shown elsewhere (Pierrehumbert 1996) stochastic constraint ranking in OT as proposed by Anttila (to appear) and others is not sufficient to capture this behavior. In general, Bybee is justified in observing that such observations lead one towards exemplar models, as proposed in the psychology literature by Nosofsky (1992) and Kruschke (1992), and in the Natural Language Processing literature in Bod (1995). In such models, access

and retrieval for structures of any type is carried out on the basis of an immense number of stored examples. Frequency information is encoded directly in the frequencies of properties of tokens in the database, and hence is implicitly updated every time a new memory is stored.

Several different methods of combining information in an exemplar approach are possible. One of the methods that Bybee considers as a viable alternative is connectionism. That is, the various exemplars are connected in a network, according to their relatedness in each of the many relevant respects. The net effect of spreading activation and inhibition gives rise to the production or perceptual analysis of individual tokens in individual cases. As argued most cogently in Dell (1996), connectionist architectures yield a generic connection between frequency and similarity which is absent in all currently competing architectures. Since any single phonological or phonetic dimension can be understood in terms of similarity in some respect, the patterns reported by Bybee are plainly amenable to a connectionist treatment. Thus, they provide telling evidence about the architecture of phonology in general.

5.   Dimensionality

In section 2, I discussed the problems posed by the granularity of phonology, in comparison with that of phonetics. Now I would like to turn to an issue which is sometimes confused with granularity, but is actually completely independent. This is the issue of dimensionality.

The vowel space provides a simple example of dimensionality. Detailed experiments on synthesis and perception of vowels have shown that vowels are very well characterized by three formants, or resonances of the vocal tract. Higher formants may be disregarded because they are not reliably excited by the glottal waveform and fail to be perceptually salient under normal listening conditions. Although a two-dimensional plot of the vowel space (F1 and F2) does much of the job and is most commonly found in phonetics texts, a more exact model also includes F3. The corresponding space could accordingly be represented by an actual three dimensional object, with particular vowel tokens represented by points (or ordered triplets), and the entire human capability for vowels represented by a three-dimensional blob. In fact, perspective drawings of such three-dimensional vowel spaces are sporadically found in phonetics articles.

The vowel space as viewed in terms of formants is a pretty low-dimensional bit of phonetics. (The number of important formants, namely three, is small.)

Cutting edge work in acoustic and and articulatory modeling uses much higher dimensional models. If we understand a "dimension" to be a degree of freedom, or a parameter which is at least roughly independently controllable, then these models involve many more than three dimensions. For example, Atal *et al.* (1978) adopt from algebraic geometry the concept of a "fiber" to organize the many different articulatory configurations which can yield the exact same vowel, as defined by the acoustics. The reader may acquire an intuitive understanding of a fiber by undertaking to articulate /a/ in two different ways: first, by retracting the tongue without moving the jaw (for example by keeping the molars in contact) second, by lowering the jaw without activating the tongue muscles. The secret to producing /a/ is to have a relatively larger cross-section in the front of the mouth than in the back, so both methods work. In fact, jaw lowering and tongue retraction can be traded off against each other in a continuous fashion, leading to a continuous family of articulatory configurations which all count as /a/. The difference between this example and the Atal *et al.* model is that Atal *et al.* use a rather finely specified acoustic tube to model the vocal tract, with the result that the fibers (or functionally equivalent configurations) emerge as trajectories through a 17–dimensional hyperspace. As a second example, the reader is invited to consider Farley (1996), which analyzes the functional behavior of a "simplified model of the larynx". This model describes 17 muscles, ligaments, and cartilages. Figures 4 through 6 of his paper display the functional consequences, in terms of 8 parameters, for contracting any of six individual muscles in the presence of various levels of contraction of the other muscles. In short, these particular figures present individual projections of a reality which is actually being represented mathematically in a 14 dimensional space.

Turning from phonetics to phonology, note that connectionist architectures are high dimensional. In contrast, generative phonology, as formulated in Chomsky and Halle (1968), is one-dimensional. The single dimension is provided by the extrinsic ordering of the phonological rules. This is the only dimension along which any kind of numbering is defined, and everything else in the model manipulates symbols logically without any concept of distance. With his contributions in experimental phonetics, Halle presumably understood that phonetics is multi-dimensional. But his model takes the phonetics to be a separate and extra-linguistic part of the overall description. Optimality Theory inherits from Chomsky and Halle (1968) the assumption that the phonology proper is one dimensional. The difference is that the dimension is provided by the constraint ranking. All the constraints fall on a line, from the lowest to the highest ranked.

Now, consider what this means for the shared goal of the Bybee and Hayes papers, which is to understand how phonetics provides a foundation for phonology. To achieve this goal, we need a method for importing phonetic patterning into phonology. For example, it is necessary to import from Farley's work the tradeoffs and synergies in articulation and acoustics which are responsible for the connection between voicing for obstruents ([+voice]) and relative f0 (L vs H tone). These are among the phonetic interactions to which Hayes alludes in his article. We want to start with a model such as Farley's, which exhibits quantitative cumulative interactions in (at least!) fourteen dimensions, and we wish to map key aspects of this model onto a phonological architecture. I assume that the connectionist architecture, being extremely high dimensional, either adopts the phonetic dimensions as such or opportunistically aggregates them if they are functionally correlated. So for Bybee, the high dimensionality of phonetics presents no problems in principle. In practice, there are of course serious challenges in working out the details.

For Hayes, the high dimensionality of phonetics poses more of a problem, because the end result (a set of constraints rank ordered under OT) has only a single dimension. Now in principle, it is always possible to map a high dimensional space onto a lower dimensional space in a one-to-one fashion, as Peano has shown (see Munkres 1975 or other introductory textbooks in topology). For example, one can quantize Farley's model with infinitessimal accuracy and map all the resulting "model points" onto some line, thus providing an infinitely exact "phonological" model of phonetics. I will not attempt to address the issue of whether this line would represent a "constraint ranking", but instead rush on to a more foundational remark. Peano mappings are not, and cannot be, both one-to-one and continuous. That is, they necessarily sacrifice the notion of a "neighborhood", or the ability to speak of what is near a point in an arbitrary direction.

This point was illustrated at the Milwaukee meeting by a modification of a method used to motivate integration by the shell method in the Northwestern University introductory calculus course. Consider a roll of paper viewed end-on. The center of the roll represents the origin. A radial direction is marked on the paper with dye. In general, two points can obviously be close either by being close along a radial direction, or close along the perpendicular (tangent) direction. Closeness in other directions can be characterized in term of radial and tangential components. A mapping of this plane onto the line which preserves adjacency in the tangent direction can be visualized by unrolling the roll of paper. Once this has been done, points which were adjacent along the radial direction are dispersed discontinuously over the paper. In short, it is not possible

to preserve continuity in both the radial and the tangential directions while mapping the plane onto the line. Only a single mapping (unrolling of a spiral) has been examined here, but the same observation holds true of any mapping whatsoever.

Therefore, we may pose the question: "To what extent are the neighborhoods defined by phonetic models preserved in phonology?" I note first that all available psycholinguistic evidence points to the overwhelming importance of lexical neighborhoods in the processing of words, in both perception and production. (See, for example, Luce *et al.* 1990; Cluff and Luce 1990; Frisch 1996). That is, data gathered from experiments on word identification, word/nonword decisions, word association tasks, and speech errors all indicate that the processing of any individual word is affected by the number and character of words which have almost the same phonological shape. Any cognitively plausible model of lexical access needs the power to refer to the similarity between any pair of words in all phonological dimensions. But results of this type do not yet get us to the traditional core problem of phonology, namely "what is a possible word?". This question may be addressed by comprehensive studies of the lexicon and/or experiments on acceptability or wellformedness.

In the discussion of table 1, I have already mentioned a case in which a deep regularity in the lexicon shows a cumulative interaction of two dimensions of the phonological description. Pierrehumbert (1994), Frisch et al (1996) and Frisch (1996) present an intensive study of a phonotactic constraint – OCP-Place in the Arabic verbal roots – which also takes up this question. The strength of the OCP-Place is found to depend on phonological similarity computed over all features, including redundant ones. That is, similarity is computed in all featural dimensions, with the result that the similarity of any two phonemes must be understood in terms of their distance in a high-dimensional space. Furthermore, there is a gradient cumulative interaction between similarity and distance, with the prohibition against oversimilar elements weakening as a function of the number of intervening elements. Berkley (1994), and Buckley (1993) have also replicated this effect for English and Tigrinya. The lexical inventories of English and Tigrinya make it possible to access the effect for more different distances than can be observed in the triconsonantal verbal roots of Arabic.

To people who are used to working with continuous mathematics, it is obvious that any attempt to recode a high-dimensional computation of similarity onto a one-dimensional constraint ranking will sacrifice the coherent treatment of neighborhoods. However, many readers of this book do not regularly use continuous

mathematics in their work, so let me attempt to walk through the nature of the problem. In Optimality Theory, the key device for describing cumulative interactions is "local conjunction". This device was introduced by Prince and Smolensky (1993) to handle situations in which violating two constraints within a domain (let's call them A and B) is worse than violating either A or B alone. Consider, for example, a hypothetical language in which the underlying representation /uj/ emerges as /wi/, despite the fact that FAITHFULNESS generally outranks the constraints ONSET and NO-CODA. That is, underlying representations which merely lack an onset emerge intact from EVAL, as do underlying representations which merely suffer from closed syllables. It is the confluence of lacking an onset and suffering from a coda which triggers the readjustment. The tableau works out as follows:

	/u j/	/w i/
A: ONSET	*	
B: NOCODA	*	
FAITHFULNESS		* *
AB: ONSET & NOCODA	* *	

Without local conjunction, the effect of violating both A and B can be no worse than the effect of violating whichever one is ranked higher. With local conjunction, the joint violation of ONSET and NO-CODA can pull ahead of FAITHFULNESS to determine the outcome. A proposal to analyze OCP-Place in this vein is recently presented in Alderete (1996). Unlike Frisch *et al.*, Alderete does not develop a complete model which allows an overall fit to a lexicon to be established. If we persevere in using local conjunction to encode the detailed effects which Frisch et al describe, we find that the constraints proliferate fantastically. For example, for a 15 dimensional featural system, we need 15 constraints to describe cases differing in a single feature, 15 times 14 for cases differing in two features, 15 times 14 times 13 for cases differing in three features ... in short, 15 factorial in all. Furthermore, to allow for statistically different behavior at four different distances (for example), we end up with 15! X 4 constraints in all. Establishing lines of "equi-strength" through this immense set of constraints in order to rank them is complicated by the fact that a constraint with two features at one distance may have the force of a constraint involving three features at some other distance. This effort is highly reminiscent of what it takes to code

floating point functions in assembly language. It encodes the cognitively real aspects of the problem extremely indirectly at best, and as such it is an exercise in logic rather than in scientific explanation.

Local conjunction is a highly restricted device. It was devised on the assumption that cumulative interactions in phonology are highly restricted, a point developed at more length in Fukazawa and Miglio (1996). If the cumulative interactions displayed by OCP-Place are typical of phonology, then the device does not have enough power to be scientifically incisive. Are such interactions typical? Plenat (1996), in a study of conditions for affixation of "-esque" in French, found that all relevant constraints interacted cumulatively. Coleman and Pierrehumbert (in press) undertook to predict the acceptability of neologisms such as "mrupation" and "glisless" using scores returned by a stochastic parser trained on the English lexicon. Comparing different scoring methods, they found that the best predictor was the overall likelihood of the form as computed from the cumulative contributions of the likelihoods of the onsets and rhymes which comprised it. This overall score was a better predictor than a score based on the single worst, or best part. As the field moves beyond exegesis of individual examples, to consider overall models of statistically valid data sets, it appears likely that more and more cases of cumulative interactions will be documented.

6.   Conclusion

Phonology is in large part built on phonetics, but it is more granular. Both Bybee and Hayes make thought-provoking attempts to show how phonetic factors are projected into the phonology, but the problem of granularity is still far from solved. Pierrehumbert, Beckman, and Ladd (1996) portray this problem as one of the biggest and most important we now face.

Bybee's paper highlights the role of the lexicon in phonology. Data of the type she describes present serious difficulties for nonlexicalist models. I agree with her conclusions about the importance of the lexicon and I suggest that they are in accordance with the mainstream of theoretical developments in linguistics over the past two decades. She is ahead of this mainstream in taking seriously the job of modeling frequency effects and in understanding that frequency is an engine of patterning.

The influence of phonetics on the ontology of phonology is now widely accepted. Bybee's and Hayes' papers raise the issue of the influence of phonetics

on the architecture of phonology, and I have taken up this question in discussion of dimensionality. I believe that the ultimate answer to phonological architecture will include the means for defining phonetic and lexical neighborhoods. Optimality Theory as presently formulated does not provide such means.

Acknowledgments

References

Alderete, J. 1997. "Dissimilation as Local Conjunction". In K. Kusumoto (ed.) *Proceedings of the North East Linguistics Society* 27:17–32. Amherst, Mass.: GLSA.

Anderson, Stephen R. (this volume)

Anttila, A. (to appear) "Deriving variation from grammar". In Hinskens, F., R. van Hout & L. Wetzels (eds.) *Variation, Change, and Phonological Theory.* John Benjamins Publishing Co.

Atal, B. S., Chang, J..J., Mathews, M. V. & Tukey, J. W. 1978. "Inversion of articulatory-to-acoustic transformations in the vocal tract by a computer-sorting technique". *Journal of the Acoustical Society of America,* 1535–1555.

Berkley, D. 1994b. "The OCP and Gradient Data." *Studies in the Linguistic Sciences* 24, 2.

Berkley, D. 1994a. "Variability in Obligatory Contour Principle Effects". *Proceedings of 30th Meeting of the Chicago Linguistics Society*, University of Chicago, Chicago.

Bod, R. 1995. *Enriching Linguistics with Statistics: Performance Models of Natural Language Processing.* ILLC Dissertation Series 1995–14. Institute for Logic, Language and Computation. Universiteit van Amsterdam, Amsterdam.

Buckley, E. 1993. "Tigrinya root consonants and the OCP". Paper presented at the 16th annual Penn Linguistics Colloquium. Philadelphia, PA.

Bybee, J. (forthcoming). "The phonology of the lexicon: evidence from lexical diffusion". In M. Barlow & S. Kemmer (eds.) *Usage-based models of language*.

Chomsky, N. & M. Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Cluff, M. S. & P. A. Luce. 1990. "Similarity Neighborhoods of Spoken Two-Syllable Words: Retroactive Effects on Multiple Activation". *Journal of Experimental Psychology: Human Perception and Performance*. 16, 551–563.

Coleman, J. & J. Pierrehumbert. 1997. "Stochastic Phonological Grammars and Acceptability". *3rd Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop*. 12 July,. 1997. Association for Computational Linguistics, Somerset NJ, 49–56.

Dell, G. 1996. "Counting, connectionism, and lexical representation: a commentary". Paper presented at the Vth Conference in Laboratory Phonology, Northwestern University, July 1996.

Edwards, J. 1996. "Lexical representations in acquisition". Paper presented at the Vth Conference in Laboratory Phonology, Northwestern University, July 1996.

Farley, G. 1996. "A biomechanical laryngeal model of voice F0 and glottal width control". *Journal of the Acoustical Society of America*. 100, 3794–3812.

Ferguson, C. & E. Newport. 1975. "Words and sound in early language acquisition: English initial consonants in the first fifty words". *Language* 51, 419–39.

Frisch, S., M. Broe & J. Pierrehumbert. 1996. "The Role of Similarity in Phonotactic Constraints". MS, Northwestern University.

Frisch, S. 1996. "Similarity and Frequency in Phonology". Ph. D. dissertation, Northwestern University.

Fukazawa, H. & V. Miglio (forthcoming). "Restricting Conjunction to Constraint Families". *Proceedings of WCCFL,* Santa Cruz CA October 1996.

Hammond, M. (this volume)

Hussain, S. & R. Nair. 1995. "Voicing and Aspiration Contrast in Hindi and Urdu". *Proceedings of the 31st meeting of the Chicago Linguistic Society, University of Chicago*, Chicago.

Keating, P. 1984. "Phonetic and phonological representation of stop consonant voicing". *Language* 60, 286–319.

Kruschke, J. 1992. "ALCOVE: an exemplar based connectionist model of category learning". *Psych. Rev.* 99, 22–44.

Lindblom, B. 1992. "Phonological units as adaptive emergents of lexical development". In C. A. Ferguson, L. Menn & C. Stoel-Gamon (eds.) *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press, 131–163.

Luce, P., D. Pisoni & S. Goldinger. 1990. "Similarity neighborhoods of spoken words". In G. Altmann (ed.) *Cognitive models of speech processing*. Cambridge: MIT Press. 148–172.

Munkres, James R. 1975. *Topology: a first course*. Englewood Cliffs, N.J.: Prentice-Hall.

Nosofsky, R. 1992. "Exemplar-based approach to relating categorization, identification, and recognition". In F. Ashby (ed.), *Multi-dimensional models of perception and cognition*. Hillsdale, N.J.: Lawrence Erlbaum.

Pierrehumbert, J. 1996. "Cooperation and Competition in the Lexicon". Paper presented at SWOT, University of California at Irvine.

Pierrehumbert, J., M. Beckman & D. R. Ladd. 1996. "Laboratory Phonology". In J. Durand & B. Laks (eds.) *Current Trends in Phonology: Models and Methods*. European Studies Research Institute, University of Salford. 535–548.

Pierrehumbert, J. 1993. "Dissimilarity in the Arabic verbal roots". *Proceedings of the 23rd meeting of the Northeastern Linguistics Society*. Amherst: GLSA Publications.

Pierrehumbert, J. 1994. "Syllable structure and word structure". In P. Keating (ed.) *Papers in Laboratory Phonology III: Phonological structure and phonetic form*. Cambridge: Cambridge University Press. 168–190.

Plenat, M. 1996. "De l'interaction des contraints: une etude de cas". In J. Durand & B. Laks, eds. *Current Trends in Phonology: Models and Methods*. ESRI/University of Salford Press. Salford, UK. 585–616.

Prince, A. & P. Smolensky. 1993. *Optimality Theory*. RUCCS-TR2. Rutgers University, New Brunswick, NJ.

Pullum, G. 1991. "Formal Linguistics meets the boojum". *The Great Eskimo Vocabulary Hoax*, University of Chicago Press, Chicago. 47–55.

Steriade, D. 1988. "Reduplication and syllable transfer in Sanskrit and elsewhere". *Phonology* 5. 1, 73–155.

Treiman, R, B. Kessler, S. Knewasser, R. Tincoff & M. Bowman. 1996. "English speakers' sensitivity to phonotactic patterns". Paper presented at the Vth Conference in Laboratory Phonology, Northwestern University, July 1996.

In the STUDIES IN LANGUAGE COMPANION SERIES (SLCS) the following volumes
have been published or are scheduled for publication:

29. ENGBERG-PEDERSEN, Elisabeth et al. (eds): *Content, Expression and Structure. Studies in Danish functional grammar.* Amsterdam/Philadelphia, 1996.
30. HUFFMAN, Alan: *The Categories of Grammar. French lui and le.* Amsterdam/Philadelphia, 1997.
31. WANNER, Leo (ed.): *Lexical Functions in Lexicography and Natural Language Processing.* Amsterdam/Philadelphia, 1996.
32. FRAJZYNGIER, Zygmunt: *Grammaticalization of the Complex Sentence. A case study in Chadic.* Amsterdam/Philadelphia, 1996.
33. VELAZQUEZ-CASTILLO, Maura: *The Grammar of Possession. Inalienability, incorporation and possessor ascension in Guaraní.* Amsterdam/Philadelphia, 1996.
34. HATAV, Galia: *The Semantics of Aspect and Modality. Evidence from English and Biblical Hebrew.* Amsterdam/Philadelphia, 1997.
35. MATSUMOTO, Yoshiko: *Noun-Modifying Constructions in Japanese. A frame semantic approach.* Amsterdam/Philadelphia, 1997.
36. KAMIO, Akio (ed.): *Directions in Functional Linguistics.* Amsterdam/Philadelphia, 1997.
37. HARVEY, Mark and Nicholas REID (eds): *Nominal Classification in Aboriginal Australia.* Amsterdam/Philadelphia, 1997.
38. HACKING, Jane F.: *Coding the Hypothetical. A Comparative Typology of Conditionals in Russian and Macedonian.* Amsterdam/Philadelphia, 1998.
39. WANNER, Leo (ed.): *Recent Trends in Meaning-Text Theory.* Amsterdam/Philadelphia, 1997.
40. BIRNER, Betty and Gregory WARD: *Information Status and Noncanonical Word Order in English.* Amsterdam/Philadelphia, 1998.
41. DARNELL, Michael, Edith MORAVSCIK, Michael NOONAN, Frederick NEWMEYER and Kathleen WHEATLY (eds): *Functionalism and Formalism in Linguistics. Volume I: General papers.* Amsterdam/Philadelphia, 1999.
42. DARNELL, Michael, Edith MORAVSCIK, Michael NOONAN, Frederick NEWMEYER and Kathleen WHEATLY (eds): *Functionalism and Formalism in Linguistics. Volume II: Case studies.* Amsterdam/Philadelphia, 1999.
43. OLBERTZ, Hella, Kees HENGEVELD and Jesús Sánchez GARCÍA (eds): *The Structure of the Lexicon in Functional Grammar.* Amsterdam/Philadelphia, 1998.
44. HANNAY, Mike and A. Machtelt BOLKESTEIN (eds): *Functional Grammar and Verbal Interaction.* 1998.
45. COLLINS, Peter and David LEE (eds): *The Clause in English. In honour of Rodney Huddleston.* n.y.p.
46. YAMAMOTO, Mutsumi: *Animacy and Reference. With case studies on English and Japanese corpora.* n.y.p.
47. BRINTON, Laurel J. and Minoji AKIMOTO (ed.): *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* n.y.p.
48. MANNEY, Linda Joyce: *Middle Voice in Modern Greek. Meaning and function of an inflectional category.* n.y.p.
49. BHAT, D.N.S.: *The Prominence of Tense, Aspect and Mood.* n.y.p.

A full list of titles published in this series is available from the publisher.