Usage-based Phonology

Joan L. Bybee
University of New Mexico

Abstract

"Usage-based Functionalism" is the term used here to designate the particular version of American Functionalism that argues that language use shapes language structure. In this paper I will examine several basic principles of functionalist theory and show how they shape a functionalist approach to phonological issues. The proposals are that performance factors not be excluded from consideration in modeling linguistic behavior; that phonetic and semantic substance be directly related to one another; that the effects of usage, particularly token and type frequency be recognized in modeling linguistic organization; that the most general phonological analysis is not the one speakers necessarily use, since local schemas take precedence over general ones; that the units of phonology are emergent; and that the real "universals" and the explanations for them are not always a part of the grammar.

When dealing with the topic of functionalism in phonology, it is important to distinguish carefully which brand of functionalism is meant, as the term has already been used to describe the theoretical stances of linguists such as André Martinet and Wolfgang Dressler. Problems with the term came to my attention about ten years ago, when Jürgen Klausenberger, in a review of my 1985 book on morphology, described it as a functionalist treatment of morphology (Klausenberger 1988). For some reason I was surprised by this label and wrote to Jürgen asking him why he categorized that work as functionalist. He wrote back saying that the book proposed to explain form in terms of function, and that made it functionalist. I could not argue with that.

A short time later I picked up *Leitmotifs in Natural Morphology* by Dressler *et al.* (1987) and found the following statement:

> The inductivist, antifunctional stand of Bybee (1985) distinguishes her work very much from ours. (p. 20)

Just when I had gotten used to the idea of being a functionalist, I found that I was also labeled an "antifunctionalist." In addition, I seemed to have been kicked out of the club of "Natural Linguists". What was going on here? Obviously there was, and still is, more than one idea about what functionalism is.

Dressler's stand is further articulated in a 1990 article in *Cognitive Linguistics*. He characterizes the basis of Natural Theories as follows:

> It is assumed that both linguistic universals and all language systems have the teleology of overcoming substantial difficulties of language performance (including storage/memorization, retrieval, evaluation) for the purpose of the two basic functions of language: the communicative and the cognitive function. (Dressler 1990:76)

In his view, languages change so that they may function better; languages are always trying to become better semiotic systems. The reason that they never quite achieve this goal is that what is optimal at one level of structure, say phonology, may create difficulties at another level, say morphology.

The difference between this type of functionalism, which is characteristic of European Functionalism, and the type I will describe in this paper, which might have the geographic designation of "West Coast Functionalism", but which I prefer to call Usage-Based Functionalism, is a matter of teleology.[1] This is best explained using the Grocery Store Analogy, as conveyed to me by Bernd Heine. If you study a lot of grocery stores at different times and different places, you will discover a basic principle of grocery stores, which is that no matter how many customers there are, nor how many check-out lines are available, all the lines tend to be of about the same length. We could conclude that there is a basic teleology of grocery stores such that they are always trying to have check-out lines of the same length. We could call this the Isometry Principle. We could further say that this is functionally motivated for the sake of the more efficient movement of customers through the store. But we would be wrong in attributing the goal-oriented behavior to the grocery store, because it is really the customers who each individually have the goal of moving through the store in the most efficient way possible. The fact that the check-out lines tend to always be the same length is a structural artifact of the goal-oriented behavior of the individual shoppers.

The type of functionalism that I will outline here does not attribute goal-oriented behavior to language systems, but rather views grammatical structure as a consequence of the way language is used.[2] A multitude of language-use events,

each of which has effective communication as its own teleology, shapes what linguists view, on the level of the whole language, as general principles of grammar.

Applications of this notion to morphosyntax are well-known. Commonly used discourse patterns become conventionalized as syntactic patterns (Givón 1979). Thus topics are grammaticized into subjects, overused pronouns become agreement markers, verbs commonly used with verb phrase complements become auxiliaries. It has been shown in many areas of grammar that recurrent patterns correspond to common discourse strategies.

But how does use create and destroy structure? What processes lead to the conventionalization of discourse patterns? It must not be forgotten that use is cycled through the cognitive storage and processing system of the language user. Thus besides studying data from the perspective of language use, we must also learn more about the properties of the cognitive system that creates both discourse and grammar.

While the theme of the current volume is the comparison of formalism and functionalism, I do not see any reason why usage-based functionalism cannot also produce formal models of synchronic linguistic systems. Cognitive grammar (Langacker 1987), construction grammar (Goldberg 1995) and network (Bybee 1985) or connectionist models of morphology (such as those reviewed in Daughtery and Seidenberg 1994) are all formalizations of usage-based grammars. The point is that usage-based functionalism is much more than a system for formalizing synchronic states; it proposes to integrate factors of language use and language change into a coherent theory that explains individual language structures and cross-linguistic generalizations. In this way, usage-based theories differ from structuralist and generative theories that assume that language structure exists independently of the way language is used.

1. Competence and Performance

Given that a usage-based theory has as its goal the understanding of both usage data and the language user, it follows that performance data should inform our understanding of the storage and organization of linguistic knowledge (see Noonan, this volume). Data on the use of language in discourse and data on the language-user as performing in real time, both of which are excluded from competence, are considered important input to the formulation of a usage-based theory. Experimental and instrumental data, to the extent that they emulate naturalistic behavior, provide important input to modeling.

Furthermore a usage-based theory should be psycholinguistically plausible. Our goal should be to model systems that move ever closer to capturing what really goes on when language is used. This is a point that is highly applicable to the modeling of phonology, since substantial phonetic and psycholinguistic data is available on phonological and morphological relations. That this point still strongly separates structural and functionalist theories is evidenced by the recent emergence of Optimality Theory, which, at least in its application to phonology, has almost no features corresponding to a plausible psycholinguistic model. (See section 9 for more on Optimality Theory.)

In recent work Mark Durie (this volume) argues that one of the main differences between usage-based functionalism and structuralist theories is the inclusion in functionalist theories of the temporal dimension, which is factored out in structuralist theories. Time figures in functionalism in at least two ways. First, functionalism recognizes that language is processed in real time, that conversation takes place in real time. The cognitive capacity, the goals and strategies used in real time are what shape the conventionalized structure of language. Second, functionalism recognizes that language changes over time, that there is no stasis, but rather a continual recreation of grammar. Diachronic change is not a great lurching from one synchronic state to another; rather the propensity for and directionality of change is an inherent part of the architecture of the system. Moreover, diachronic change plays an indispensable role in the formulation of the explanations for linguistic patterns. I will have more to say about diachronic change as we proceed.

2. Substance vs. Structure

Chomsky and Halle (1968) introduce a distinction between substantive and formal universals, where "substantive" refers to the inherent content of phonetic categories, such as nasality, voicing, etc. and "formal" refers to the structural arrangement of features and segments, and the formal properties of rules and their interactions. Phonetic content is substantive because there is a physical correlate that can be described in terms that are independent of language — movements of articulators, properties of sound waves, and so on. Categorization of surface segments into phonemes, generalizations over patterns of alternations and phonotactic regularities, though they may be based on phonetics in some sense, are conventionalized, language-specific and purely internal to language.

Semantic content is also substantive in the sense that it relates directly to properties of reality as perceived by human beings. Semantic features are part of a linguistic system, but their content links to features of the world independent of language. On the other hand, morphosyntactic regularities of distribution may be based on semantics, but to the extent that they are conventionalized, they are purely internal to language and thus are structural rather than substantive.

Emergentist and connectionist views of language take substance (or the perception and memory of experience with substance) to be directly represented, while structure is considered emergent from the way substance is categorized in storage, which in turn is based on patterns of actual language use. Under this view, phonological and morphosyntactic regularities are emergent. This means that such patterns are not basic but a secondary result of aspects of speaking and thinking; they are not necessarily categorical, symmetrical or economical, but vary according to the nature of the substance involved, and the demands of communication.

Phonology, the abstract patterning of sounds in the service of conveying meanings, is emergent. Because phonology associates with both of the substantive ends of language — phonetics and semantics — its study requires an understanding of the principles of both. Since these two types of substance play very different roles in phonological phenomena, it is important to know in any given case to what extent a phonological phenomenon is governed by one or the other. Natural theories (Stampe 1973; Hooper 1976a) have always made the distinction between automatic, phonetically-motivated processes and rules with lexical or morphological conditioning, recognizing that these two types of patterns have very different properties. However, this distinction only makes sense as it is applied to surface-level phenomena where there can be direct empirical confirmation of the substantive category a pattern belongs to. It is meaningless to make this distinction where rule ordering or level ordering allows rules to be formulated as exceptionless despite having surface exceptions (Hooper 1976a).

A substantive pursuit consistent with Usage-Based Functionalism is the study of the properties of these two groups of patterns. The understanding of processes with phonetic conditioning naturally must rely on phonetics, that is, an understanding of the dynamics of production and perception mechanisms. While a lot can be learned from typological studies of possible sound changes and phonetic processes, the explanation for the range of possibilities will be grounded in phonetics. Conditioning for alternations from the non-phonetic domains has been less studied and consequently is less well understood. What are the factors that encourage or inhibit a phonetic process in one word but not another? What

are the conditions under which an alternation comes to be associated with a morphological category? How are lexical classes of items with alternations organized? Questions concerning the interaction of phonetic processes with grammatical and lexical categories are best approached from a diachronic perspective.[3] We can think of phonological alternations as having life cycles or unidirectional paths of change (just as in grammaticization) with particular mechanisms of change applicable at different stages: alternations begin for purely phonetic reasons and gradually become conventionalized as part of the sound / meaning correspondence of the language, in the process interacting more and more with the lexicon and morphology (Hooper 1976a; Klausenberger 1979; Tranel 1981; Vennemann 1972). How this process takes place is an intriguing issue that has been studied more by historical linguists and variationists than by synchronic phonologists, even though this dynamic aspect of phonology holds the key to our greater understanding of how phonological patterns are internalized cognitively. I believe that we will be able to formulate stronger and more revealing universals of phonology by proposing universals of change than by proposing synchronic universals (just as we can for grammar, see Bybee, Perkins and Pagliuca 1994). Every diachronic change contributes to a synchronic state, but because diachronic changes can be combined in so many ways, the number of different synchronic states is much greater than the number of diachronic changes. The next section discusses some aspects of the diachronic trajectory for phonological alternations.

3. The Evolution of Phonological Alternations

While everyone probably agrees that phonological patterns eventually become lexicalized and in some cases morphologized, there is considerable disagreement about the timing and the mechanisms by which the processes of lexicalization and morphologization take place. There is now considerable evidence that lexical and morphological factors can figure very early in the development of phonetic processes. This evidence cannot be accounted for in structuralist or modular theories of language, and has therefore, to the extent that it has been recognized at all, been swept under the rug, usually with the label of "low-level phonetic detail."

The theoretical problem is as follows: structuralists proposed that allophones had to be predictable in purely phonological terms and only phonemes could participate in morphologically conditioned alternations. Thus the phonetic, phonemic and morphophonemic levels were strictly separated. A very similar

result is achieved in Lexical Phonology with the Structure Preservation constraint, which requires that the output of lexical rules consist entirely of segments or feature combinations that are possible in underlying forms. If a rule introduces or refers to a noncontrastive feature, it can only apply post-lexically (Kiparksy 1982). These constraints capture a very important universal tendency: that alternations that are involved with the lexicon or morphology are alternations among phonemes, while those with only phonetic conditioning may involve subphonemic features.

However, these constraints do not explain this property of language and they fail in certain cases where subphonemic features are partially conditioned by lexical and morphological features. Such situations arise as language changes: in order for an alternation to become lexicalized or morphologized, there must be some intermediate stage in which a feature that appears to be noncontrastive has actually achieved lexical or morphological status. By studying such cases as dynamic situations that are changing over time, we will begin to understand why and how such changes occur and why languages are in general structured in accordance with the constraints described above. That is, diachronic data can lead both to an understanding of the exceptions and an explanation for the constraints.

A well-known case is the alternation of [x] and [ç] in German, where the palatal fricative occurs after front vowels and /n/, /r/ and /l/, and the velar occurs elsewhere. The diminutive suffix *-chen*, however, always contains the palatal fricative, even when the preceding noun ends in a back vowel, as in *Tauchen* [ç] 'small rope' and *Pfauchen* [ç] 'little peacock'. (These form minimal pairs with the verbs *tauchen* 'to dive' and *fauchen* 'to spit', which have the velar fricative.) The traditional solution to this problem (Moulton 1947; Leopold 1948) is to predict the [ç] from a boundary (or juncture), which avoids recognizing it as a phoneme. It also seems important to generativists to keep [ç] out of the lexical inventory, so this type of solution is continued in generative phonology. Since it violates Structure Preservation, Hall (1989) argues for a relaxation of Structure Preservation in this case (but see Macfarland and Pierrehumbert 1991 for another proposal).

My view is that it is preferable to regard this bit of messiness as revealing something that is hidden in the more usual phonemic relations: what factors operate to create a new phoneme. First, note that the phonetic distance between the [x] and the [ç] is greater than that necessitated by the position of the preceding vowel, and that a palatal and velar fricative do contrast in some languages. Second, the palatal may occur outside the phonetic conditioning

environment of a preceding front sound, in the diminutive suffix, as we have noted, but also after the uvular [R]. Presumably, the /r/ was alveolar (like /n/ and /l/) at the time the [ç] came to appear consistently after it. It is telling that when the /r/ became uvular, the [ç] did not also revert to [x]. Finally, the palatal has gained enough autonomy to be used in certain loan words in initial position or even after a back vowel, because it resembles the consonant in the source language more than any other German consonant: *China* [ç] and *Photochemie* [ç].

In seeking to tidy up language, the generative tendency is to try to get [ç] back in line as an allophone derived from /x/. However, if we do that, we are missing something very interesting in this situation. The palatal fricative has several properties that distinguish it from a normal allophone and these properties have presumably developed in the same time frame and are related to one another. Can we seek a cause and effect relation among them, or are all of these properties intertwined?

Although there has not been sufficient study of such cases in the midst of change, my hypothesis is that the leading cause in such cases is the phonetic substance: phonetic categorization (which will be dealt with a little more below) depends upon phonetic similarity; two sounds must be highly similar to be considered members of the same category. It may be that at some point [x] and [ç] diverged too much to be assigned to the same category and a new category was created for the palatal. Thereafter, any changes in its environment would leave it unaffected. In such cases, the nature of the substance determines the assignment of structure.

A causal relation that we can reject is that the loss of the phonetic conditioning led to [ç] becoming phonemic. Certainly the loss of the phonetic environment forces an analysis of [ç] as phonemic, but it is not the cause of the change. This can be argued on logical grounds: if [ç] were noncontrastive, the loss of the phonetic environment would cause it to revert to [x]. But there is also empirical evidence showing morphologization (and thus the attainment of contrastive status) where phonetic conditioning is still present. The following case illustrates this point and also argues that morphologization occurs much earlier than is usually supposed. This is not a new example (it appears in Hooper 1976a) but I fear that its import has not been sufficiently appreciated.

Throughout much of the Spanish-speaking world, a gradual, variable deletion of syllable-final /s/ is underway. This phonetically-motivated process has consequences for the morphology of the language, since a word-final /s/ in most cases is the signal of plural in nouns, determiners and adjectives, and the second

person singular in verbs. It is noteworthy that in most dialects the deletion proceeds despite the potential loss of morphological information. Studies of the change in progress show that potential ambiguity is not usually a deterrent to deletion. However, there is one dialect in which a new phonological contrast arises in noun plurals even before the traces of final /s/ are completely gone.

In most dialects of Spanish there is noncontrastive opening or laxing of vowels in closed syllables and this process affects the vowel before a syllable-final /s/. In the Granada dialect of Spain, where syllable-final [s] has been weakened to [h] or deleted, this vowel laxing has become the signal for plural in nouns and adjectives (Alonso, Vicente and Canellada de Zamora 1950).[4] In fact, as Alonso *et al.* show, a vowel harmony system has developed in which all the vowels of the plural noun or adjective are lax, while all the vowels of the singular are tense, whether or not they occur in open or closed syllables.

(1)	*Orthographic*	*singular*	*plural*	
	pedazo	[peðaθo]	[pẹ̣ðạ̣θọ̣]	'piece'
	alto	[alto]	[ạ̣ltọ̣]	'tall'
	cabeza	[kaβeθa]	[kạ̣βẹ̣θạ̣]	'head'
	selva	[selva]	[sẹ̣lvạ̣]	'forest'
	lobo	[loβo]	[lọ̣βọ̣^{h}]	'wolf'
	tonto	[tonto]	[tọ̣ntọ̣]	'stupid'
	piso	[piso]	[pị̣sọ̣^{h}]	'floor'
	fin	[fiŋ]	[fị̣nẹ̣^{h}]	'end'
	grupo	[grupo]	[gṛ̣upọ̣^{h}]	'group'

The examples in (1) are shown as transcribed in Alonso *et al.* It is noteworthy that they transcribe final aspiration in plurals about half the time, meaning that the phonetic remains of the /s/ are still present in the language. They also observe that the phonetic distance between tense and lax vowels is greater in this dialect than in neighboring ones.

The evidence that the lax vowels signal plural and are not just in phonological agreement with the final vowel is that in words such as the days of the week, which end in /s/ in both the singular and plural, the vowel laxing occurs only in the plural:

(2)	*el martes*	[marte]	*los martes*	[mạ̣rtẹ̣]
	'Tuesday'		'on Tuesdays'	

I want to draw attention to two points that follow from the fact that morphologization has occurred before the /s/ was entirely deleted. First, the reanalysis of

laxing as morphologically motivated rather than phonologically motivated occurred at a time when the lax vowel was still predictable phonologically. That is, the speakers chose a morphological analysis over a phonological one while both were still possible. Second, this example shows the very early involvement of the morphology in the development of fine-grained, phonetic alternations.

If there were a constraint on grammar that lexical representations and the output of lexical or morphological rules contain only contrastive segments, there would be no way to have lax vowels as the output of the morphologized harmony rule. Indeed, there would be no way for lax vowels to become contrastive in this dialect. Similarly, there would be no way for [ç] to become contrastive in German, or for any new phoneme to become established in a language. The only possible conclusion in the face of this type of data (which is actually quite common) is that at least in some cases noncontrastive features and segments must appear in lexical representations and in morphological patterns.

4. Lexical Effects in Variation

A related issue is the extent to which variable processes affect particular lexical items in differential ways. For example, it is well-known that the process of t/d-deletion in American English has for some speakers totally affected some high frequency words, such as *just, went* and *and*. This means that the lexical representations of these words have been restructured as the result of a very low-level rule. Furthermore, there is considerable evidence that shows that sound change or variable processes affect words gradually and at different rates, with a significant variable being the frequency of use of the word. High frequency words have been shown to undergo many sound changes at a faster rate than low frequency words: for example, in vowel reduction and deletion in English (Fidelholtz 1975; Hooper 1976b), the raising of /a/ to /o/ before nasals in Old English (Phillips 1984), in various changes in Ethiopian languages (Leslau 1969), in the weakening of stops in American English and vowel change in the Cologne dialect of German (Johnson 1983), in ongoing vowel changes in San Francisco English (Moonwomon 1992), and in tensing of short *a* in Philadelphia (Labov 1994:506–7).

For the case of t/d-deletion in American English, I have shown that across 2 000 tokens of word-final /t/ and /d/, there is a significantly higher rate of deletion in high frequency words than in low frequency words (Bybee 1997). Table 1 shows the difference in the deletion rate for high frequency words, those with a frequency in Francis and Kučera (1982) of 35 or more per million, versus low frequency words, those with a frequency of less than 35 per million. A similar frequency effect is also present among regular past tense verb forms and among the double-marked pasts, such as *told* and *felt* (see Bybee 1997).

Table 1. *Rate of* ***t/d*** *Deletion for Entire Corpus by Word Frequency*

	deletion	*non-deletion*	*% deletion*
high frequency	898	752	54.4%
low frequency	137	262	34.3%

Chi-squared: 41.67, $p < .001$, $df = 1$[5]

My proposal for accounting for this frequency effect is to conceive of variable processes as applying in real time, opportunistically, each time a word is used (a proposal based on Moonwomon 1992). Words that are used frequently in social contexts where sufficient compression and reduction occur to produce perceived deletion will have a more reduced /t/ or /d/ than words that are less frequently used in such contexts.[6] Each token of use of an item affects its memory representation. Since tokens of use vary, the stored representation must include a range of variation. As words slowly and gradually reduce in production, the center of the range of variation gradually shifts.

Thus in order to account for this frequency effect on particular lexical items, it is necessary to have a lexicon in which considerable phonetic detail and ranges of phonetic variation are represented with each word or phrase. Not only do lexical representations have to be fully specified and represented in concrete phonetic units, these units cannot be an idealized systematic phonetic set of units, but rather must represent in some realistic way the range of variation occurring in the individual pronunciations that are constantly being mapped onto the existing representations.

Experiments on perceptual categorization lead to the same conclusions. In a series of experiments, Joanne Miller and her colleagues have demonstrated that phonetic categories have rich internal structure and that subjects can judge how well a given stimulus fits into a phonetic category (Miller 1994). Phonetic categories have a prototype structure based on actual tokens. Nosofsky (1988) has demonstrated that the frequency of a stimulus helps to determine the structure of a category, with stimuli identical or similar to high frequency exemplars being judged as highly typical of a category. Thus the most-used

variants determine the prototype of the category. Nosofsky accounts for these facts by assuming "that people learn categories by storing individual exemplars in memory" (1988:62). In a linguistic model the relevant exemplars are words or phrases, which are themselves categorized both phonetically and semantically, and which in turn consist of smaller units (morphemes, syllables, segments) which are also categorized by association with similar configurations in other words or phrases. Thus I am claiming that ranges of variation in the phonetics of individual words are a part of the language user's knowledge about these words, just as much as the detailed knowledge about the linguistic and non-linguistic contexts in which the word has been used. Abstractions and generalizations over these detailed representations are expressed in schemas that emerge from these representations, as we will see in sections 7 and 8.

Miller also observes that the identification of a prototype may depend on context and there can be different prototypes for different contexts (including e.g. rate of speaking). Since assignment to a category depends upon phonetic similarity, one could imagine a situation in which two context-dependent prototypes (such as [x] and [ç] in German) gradually move away from each other until they are no longer exemplars of the same category. The result is the gradual creation of a new category, that is, a new phoneme.

Examples of lexically-determined subphonemic variation, such as that due to frequency, show that we must recognize the interaction of even low-level phonetic phenomena with the meaningful end of grammar or lexicon. Form and meaning should not be separated by multiple layers of structure, whether they are called levels or modules or components. My point in this section, then, is that substance is directly related to substance in a usage-based model. The phonetic percepts that are input and output of the grammar are related in a very direct fashion to the meanings they convey and the contexts in which they occur. Phonological processes become entwined with the meaningful aspects of grammar much earlier than is usually recognized.

5. Effects of Usage

In the previous section some effects of token frequency were mentioned in connection with the argument that even very "low-level" phonological phenomena have effects on particular lexical items. In this section I will outline the two prominent effects of token frequency, discuss the role of type frequency, and explain their interactions, with the goal of showing that language use has a significant impact on phonological and morphological structure.

The two effects of token frequency (the frequency of individual items) seem contradictory since one seems to accelerate change while the other seems to promote conservative behavior. I will argue that the reductive effect of frequency is related to the automation of production, and that the conserving effect of high token frequency affects storage and access to linguistic strings. Both of these effects are related to the fact that language is a conventionalized and even ritualistic cultural object (Haiman 1994).

One effect of high token frequency is the reductive effect, which was discussed in the preceding section. Words and phrases that are used more often undergo compression and reduction as part of the move to automate speech: boundaries are obscured and segments and syllables may disappear into the mass of co-articulated gestures. In the previous section we saw evidence for the accelerated reduction of frequent words. Further study of frequent words shows that they reduce more in the contexts in which they occur most frequently, because in these contexts they become just a part of a larger processing unit (Anderson 1993; Boyland 1996).

Consider the reduction of *don't* in American English. Kaisse (1985) has approached this variable reduction from a syntactic point of view, trying to discover the syntactic configuration that leads to the reduction of *don't*. Even in a syntactic approach, however, it is necessary to acknowledge that reduction of the vowel of *don't* to schwa is possible only with pronoun subjects and a small set of verbs such as *know, want, care, mean, like*, etc. Furthermore, looking at the variation of *don't* in actual conversation, one discovers that *don't* reduces the most in the contexts in which it occurs most often. Scheibman (to appear) and Bybee and Scheibman (1997) find that in 138 occurrences of *don't* in spontaneous conversation, reduction to schwa occurs 51 times, and of these, 50 cases occur with *I* as the subject. *I* is also the most common word to precede *don't*, occurring 88 times (63% of the cases). Furthermore, reduction to schwa occurs only with the verbs that are most frequently used with *I don't*: *know, think, have, want, like, mean, care* and *feel*. The one case of reduced *don't* that occurred with a word preceding *don't* other than *I*, was in the phrase *why don't you*, used to make a suggestion, as in *why don't you sit down*.

The evidence is very clear: the reduction of *don't* occurs in high frequency phrases that have become established as processing units. A chunk of speech, such as *I don't know*, has become an autonomous storage and processing unit, not because it is in some way idiosyncratic, but simply because it is frequently used, and can be processed more efficiently as a single unit than as a concatenation of

four units (Anderson 1993; Boyland 1996). Processing it as a unit leads to reduction: in this case the stress on the second element, *don't*, is reduced and consequently the vowel undergoes reduction as well. The /d/ also reduces to a flap and deletes in some cases (again, only in the most frequent contexts).

The effects of repetition are not restricted to the phonetic substance; the semantic substance is also subject to reduction of its significance and depletion of its force, enabling a form to acquire grammatical or discourse functions, as in the process of grammaticization, exemplified here by the phrase *I don't know* which takes on the discourse function of modifying a speaker's stance towards what s/he has said (Scheibman, to appear). Individual units within a phrase lose their semantic autonomy and phrases lose their internal structure because of this reduction process. With Haiman (1994), I would argue that repetition is driving the automatization that leads to phonetic reduction, semantic reduction and the coalescence of multiple units into a single unit.

The other effect of high token frequency is lexical strength (Bybee 1985) or entrenchment (as Langacker 1987 calls it). While rituals are reducing in form and losing their original meaning, they are also, by repetition, becoming more entrenched and conventionalized as part of the culture. So, too, with linguistic forms that are of high frequency. Their greater entrenchment leads to greater strength of storage in memory—they are thus easier to access and more resistant to change under the pressure of patterns emerging from other forms (Bybee 1985).

Lexical strength manifests itself in several ways. One obvious effect of high token frequency is the maintenance of morphological alternations that would otherwise be leveled. A high level of use creates a stronger representation for an irregular past such as *kept*, which makes it much less likely to regularize to *keeped* than a lower frequency verb of the same pattern, e.g. *wept*, which does regularize to *weeped* (Bybee 1985). Regularization takes place if the regular pattern is easier to access than the irregular form, as when it is of low frequency, and thus has a weak representation. High frequency irregular forms are so entrenched that they are unlikely to ever be regularized (though they may be replaced entirely), leading to the universal pattern that morphological irregularity is always situated in the most frequent nouns, verbs and adjectives of a language.

Related to the stronger stored representations of frequent forms is their more autonomous status (Bybee 1985). Morphologically related words are stored with connection to their relatives: *happy, happiness, unhappy; drive, driving, driven, drove* are stored close to one another, in neighborhoods that involve both phonological and semantic similarity (Pisoni, Nusbaum, Luce and Slowiaczek 1985), probably organized around the strongest member, *happy* and *drive* in these cases (Lukatela, Gligorijevic, Kostic and Turvey 1980). But morphologically complex words that are very frequent are less dependent upon related forms and tend to move away from their relatives both semantically and phonologically. For example, *despair, desperate*, and *devout, devotion* are derivational words that are, in my own opinion, not especially closely related, and in both cases the "derived" form is at least as frequent as the "base" form.[7] Thus the "derived" form is capable of being autonomous from its etymological base. In inflection such splits are much less common, but we do see them in the process that creates suppletion: *went* split off from its base form *wend* and joined the paradigm of *go*. This shift must have been made possible with an extreme increase for *went*.

Both the reducing and strengthening effects of high token frequency require direct representation in the lexicon. The relatively faster reduction in certain lexical items caused by general processes, such as t/d-deletion, is part of the phonetic knowledge that the user has about each word. The greater entrenchment that comes from the reinforcement of repeated use results in a higher level of activation, even at resting levels. It also results in weaker connections among related forms of high frequency. Thus the lexicon is neither uniform in the units of representation nor in the status of the lexical entries themselves. The lexicon is highly affected by language use.

6. Type Frequency and Productivity

The patterns or schemas that emerge from the represented units are also affected by language use. The degree of productivity of a morphological pattern corresponds quite closely to type frequency, that is, the number of lexical items participating in the pattern. Studies on English by Moder (1992) and Wang and Derwing (1994), and on Hausa by Lobben (1991) have shown that the greater number of lexical items participating in a schema the greater its likelihood to apply to new items. Schemas that range over lots of items are themselves stronger and easier to access. Patterns involving only a few items (fewer than five) usually do not constitute schemas at all.

These facts call into question the description of some relic patterns applying to only a few items in terms of rules, lexical or otherwise. What constitutes a viable schema is an empirical matter, that can be determined on the basis of speakers' reactions to novel forms (Bybee and Moder 1983; Wang and Derwing

1994; Lobben 1991). Many earlier assumptions about "rules" need to revised. First, just because we can identify a distributional pattern does not mean that speakers have organized their lexicons in terms of this pattern — many patterns reflect defunct diachronic processes; second, even if a word fits a viable schema it does not necessarily participate in the schema — words of high frequency can be independent of other generalizations in the lexicon just as they are more independent of related forms (Bybee 1985; Moder 1992); third, schemas are not necessarily formulated in the most general way possible — not all segments or morphological categories that could be subsumed under a single schema actually participate. That is, once out of the domain of the phonetic generalization, it is not the case that speakers always go for the most general and abstract schema (see the next section, as well as Aske 1990; Bybee and Pardo 1981; Lobben 1991; Wang and Derwing 1994).

These points could be illustrated with any number of examples, but I will start with a Spanish example for which experimental evidence is available. In all Third Conjugation verbs with the mid vowels /e/ and /o/ the stem vowel changes to the corresponding high vowel in the Third Person Preterite:

(3)	*mentir*	'to tell a lie'	*dormir*	'to sleep'
	mentí	mentimos	dormí	dormimos
	mentiste		dormiste	
	mintió	mintieron	durmió	durmieron

Clearly these parallel alternations should be handled with one generative rule affecting both front and back vowels (Harris 1969). However, there is a marked asymmetry in the type frequency of the alternations with front vs. back vowels: while more than 60 verbs have the front vowel alternation, only two verbs with back vowels have this alternation (*dormir* 'to sleep' and *morir* 'to die'). Thus it is not surprising that in a nonce-probe task, more responses show a generalization of the front vowel alternation than the back vowel alternation. Given the nonce forms *rentir* and *sornir* twelve out of twenty-two subjects gave the 3s Preterite *rintió* and only one out of twenty-two gave the corresponding back-vowel form *surnió* (Bybee and Pardo 1981:943–946).[8] Thus the productivity of the front vowel alternation does not extend to the back vowels, suggesting that speakers do not include *dormir* and *morir* in the schema for the front vowel alternation.

Two conclusions can be drawn from this result. First, phonological generalizations that are significant for phonetically motivated processes are not necessarily retained in morphologized patterns; that is, the organizing principles for morphology are distinct from those found in phonology. Second, high frequency verbs such as *dormir* and *morir* can be learned and accessed independently and do not necessarily participate in a schema, even if their forms fit the pattern described in the schema. Patterns of usage are more important in determining the nature of storage and access than distributional patterns.

Consider another example that appears in a popular phonology textbook, Kenstowicz (1994). This example involves the so-called alternation of the words *damn* with *damnation*, *condemn* with *condemnation* and *hymn* with *hymnal*. In this particular case the "rule" involved expresses a valid generalization about English — that no word ends in two nasal consonants. But the use of this "rule" to relate the three word pairs runs into multiple problems. First, the pattern involves too few items to constitute a set that would support a productive schema. Second, the "rule" is formulated to delete an /n/ in the base form. In each case the base form is the higher frequency form.[9] It is not realistic to suppose that the lexical representation for *damn*, *hymn*, and *condemn* end in any consonant other than /m/. These words would have been acquired and firmly established before their derivatives, and it is not reasonable to suppose that a person who has been saying and hearing [dæm] will decide that the word actually ends in two nasals upon hearing *damnation* for the first time and realizing that it is associated with *damn*. Finally, the problem that plagues the Lexical Phonology treatment of these words is that the /n/ does not show up in *damning*, *damned*, *condemning*, *condemned* or *hymns*, which are instead taken directly from the base as pronounced, even though the /n/ is not word-final.

It is not my goal to criticize any particular analysis of these words, but merely to point out that taking usage into account provides a very different focus of interest. Instead of turning up all possible alternations that a clever linguist can discover and describing them through sets of rules or the interaction of constraints, a usage-based treatment would ask how speakers conceptualize and process relations among words, based on their experience with them.

7. Redundancy and Generalization

As Langacker (1987) has pointed out, a cognitively plausible phonological theory must reject the strict distinction, so important in structuralist and generativist theories, between idiosyncratic and thus lexically-listed properties of linguistic units and predictable, redundant or rule-derived properties. We need instead a rich and highly redundant means of storage, highly responsive to ongoing

experience with language, from which generalizations (called schemas) of various levels of abstractness emerge. A few arguments are offered here:

The recent literature on the psychology of categorization makes it clear that human beings do not divide the features that characterize categories into two types: contrastive and predictable. Rather exemplars are categorized according to the number and type of features they share with the central members of the category. The status of the features depends upon the frequency with which they occur in members of the category in our experience, and upon their cue validity (how well they distinguish members of this category from members of other similar categories). Even redundant features may be very important to the categorization of a token as a member of the category.

Phonetic studies have confirmed the importance of structurally redundant features to perception: in English, vowel length provides an important cue to final consonant voicing, even though it is largely predictable by rule. Vowel transitions provide the cues to identification of the point of articulation of consonants. Moreover, it is often multiple, co-occurring cues that yield accurate identification, showing that we cannot choose one feature as distinctive and weed all others out of lexical representation (Ohala and Ohala 1994).

An early assumption in generative phonology seems to have been that neural storage space was limited and that simple storage and complex computation was highly efficient. This assumption is now known to be incorrect. The amount of neural storage space is not an issue. This does not mean that there is no parsimony in storage; indeed, categorization itself is the organizing feature that creates economy in storage: similar items are stored together and generalization over these items is manifest at many levels.

Another traditional assumption in generative phonology, which continues into the more recent constraint-based theories, is that a linguist's goal should be achieving generality at the level of theory and at the level of language-particular analyses (Prince and Smolensky 1993:4). Usage-based theories also aim at generality in the postulation of theoretical constructs, but they do not assume that particular linguistic analyses are maximally general. In fact, there is considerable evidence to support the proposal that speakers form overlapping local generalizations which are accessed more readily than the more abstract, general schemas that range over them (Aske 1990; Langacker 1995; Moder 1992; Wang and Derwing 1994). Thus instead of seeking the most general analysis for any set of data, the research program seeks evidence beyond distribution, i.e., diachronic or experimental evidence concerning the nature and range of generalizations that speakers are able to apply. The next section briefly discusses some of this evidence.

8. Local vs. General Schemas

An interesting study by Aske (1990) showed that in some cases Spanish speakers tended to choose a more specific pattern of stress assignment over a more general one. In Spanish 95% of nouns and adjectives can be described by the general rule that those ending in consonants have final stress, while those ending in vowels have penultimate stress. But there is a significant cluster of nouns and adjectives ending in *-en* that have penultimate stress: *órden* 'order', *orígen* 'origin', *imágen* 'image', etc., which must be considered exceptions to the general rule. However, in an experiment, Spanish speakers read a majority of novel words that ended in *-en* with penultimate stress, in violation of the general rule, but in conformity with a pattern specific to words ending in *-en*. Words in ending in *-n* with vowels other than *e* (e.g. *-in*, *-an*, *-on* or *-un*) were given final stress. This evidence suggests that a cluster or gang of similar words can create a local schema that will take precedence over a more general one.

Wang and Derwing (1994) show experimentally that different exemplars of the English vowel shift are more or less productive according to the morphological pattern tested. For nominalizations of nonce forms with *-ity*, the vowels most often produced were /ɔ/, /ɪ/, and /æ/ (in that order); for past tense, the most-used vowels were /æ/, /ow/ and /ɔ/; and for plurals the most used vowels were /iy/ and /ay/. While /æ/ and /ɔ/ were used in nominalizations and past tense, their ordering in the group of most-used vowels was different. Rather than there being one very general rule for English vowel-shift alternations, it appears that each morphological pattern has its own set of productivity patterns.

Wang and Derwing also show that the productivity of the vowels in each pattern can be directly related to the type frequency of those vowels in the pattern in actually-occurring English words. They further argue that it is the vowel in the morphologically-complex word that is productive, not the relation between an "input" vowel and an "output" vowel, as a wide variety of input vowels could elicit the most popular output vowels. Thus the generalizations are product-oriented: they are expressed in generalizations over the morphologically-complex word, not as rules that turn one vowel into another (see also Bybee and Moder 1983).

This study illustrates the point that for alternations associated with grammatical morphology, speakers do not construct a single generalization based on phonological features. The experiment shows that speakers do not have a single vowel shift rule for all vowels, nor do they generalize across all morphological conditions. Rather, the data is consistent with the postulation of many local

schemas of highly similar words, such as the nominalizations *oddity, commodity, velocity, mediocrity;* the past tenses *sang, rang, hang;* and the plurals *feet, teeth, geese.* Productivity does not depend upon the general pattern, for if it did, all vowel alternations would be equally productive. Rather it depends upon how many words are in each of these local clusters.

In contrast to generative theories which avoid redundant expression of generalizations, Langacker (1995) argues that schemas are maximally redundant: the same items may be encompassed by schemas of differing levels of generality, ranging from those that generalize over a single lexical item up to those that describe regularities in hundreds or thousands of items. It follows from this that there need not be a unique analysis for any given data set, but that there may be various ways of generalizing over the same material. It can be determined experimentally which schemas are stronger and therefore easier to access, and what properties of linguistic items are important for determining the strength and level of generality of a schema.

While there is no priority set on parsimony of linguistic analysis, the actual mechanisms proposed for a usage-based model are of a very general nature and take the same form for morphology, lexicon and syntax. Thus the proposal by Pinker (1991) and his colleagues for two separate mechanisms for the processing of morphology is rejected, since the differences they note between "regular" and "irregular" morphology can be entirely attributed to usage: high type frequency produces a high level of entrenchment for morphological schemas and in extreme cases, such as the regular English Past tense, a schema can seem almost independent of particular lexical items (Bybee 1995).

9. Units of Phonology

In network models, internal structure is emergent — it is based on the network of connections built up among stored units. The stored units are pronounceable linguistic forms — words or phrases stored as clusters of surface variants organized into clusters of related words. One recent suggestion for the representation of phonological properties of lexical items is that they are represented as gestural scores that characterize the activity of the vocal tract (Browman and Goldstein 1991), and it seems reasonable to suppose that an associated perceptual image may also be part of the representation.

Units such as syllables and segments emerge from the inherent nature of the organization of gestures for articulation. Browman and Goldstein (1995:20) argue

that "syllable structure is a characteristic pattern of coordination among gestures." Once the nature and timing of gestural coordination is described, in effect, syllabic structure has also been described. Several recent studies have also argued for the emergent nature of segments. Ohala (1992) argues that the temporal coordination of certain gestures enhances their acoustic effect and thus provides an evolutionary impetus for the development of segments. Lindblom, MacNeilage and Studdert-Kennedy (1983) rated seven onsets with complete closure and nineteen steady state formant sequences (for a total of 133 possible syllables) for acoustic discriminability and articulatory ease and in a series of simulations found that the fifteen syllables in (4) emerge as the optimal syllables:

(4)

bi	bɛ	ba	bo	bu
di	dɛ	da	do	du
ɟi	ɟɛ	ga	go	gi

([ɟ] is a palatal stop.)

Among the interesting properties of this set is that out of the seven points of articulation tested, four emerge as optimal. These include the three points of articulation for stops most common in the languages of the world, and the palatal stops, which is a common variant of the velar in just the contexts in which it emerged as optimal. In addition, the five cross-linguistically most common vowels emerge in this simulation. Rather than fifteen syllables all containing different onsets and steady-state offsets, the same onsets and offsets are repeated across the fifteen, leading to a pattern from which a small set of consonants and vowels emerge.

Given that lexical storage units — words and frequent phrases — are gestural scores, we may ask if there is correlate of segment to be found in the continuous representation? A true one-to-one relationship between segments and part of the score is not to be expected, given the well-known problems with the notion of segment. However, what is present in the score are certain points of temporal coordination involving (for consonants) the achievement of the target, the beginning of movement away from the target, or occasionally onset of movement toward the target (Browman and Goldstein 1992). These phasing points among the gestures of the independent articulators correspond in a very rough way to the areas in the speech stream that we tend to identify with alphabetic symbols designating consonants. The more steady state portions of the score are identified as vowels. In this view, consonants and vowels are derivative of gestural coordination.

Moreover, recurring sets of gestural phasing can be identified in independent words or phrases. Lindblom (1992) argues that anatomically and temporally identical control functions may be stored only once. This means that parts of gestural scores of a language that are identical (or highly similar) will be linked to one another as representing the same motoric pattern. From such repeated sets of coordinated gestures, a "segment" inventory can be derived. The re-use of the same sets of gestures in the same temporal configuration is necessary if a child (or a language) is to acquire a large vocabulary (Lindblom 1992; Studdert-Kennedy 1983, 1987).

In the conception of the lexicon I am advocating here, there is a set of highly entrenched gestures and gestural configurations that are used and re-used in constructing the words of a language, but there is no reason to exclude lower frequency configurations, such as that needed for English [ʒ], or that needed for sounds with restricted distributions, such as English [ŋ], or new configurations that arise through the reduction and retiming of old configurations, such as new consonant clusters. It also makes it unnecessary to make arbitrary decisions about the locus of contrast in transitional cases such as the English case of vowel length differences before voiced and voiceless consonants. Since both vowel length differences and glottal opening can be represented in the gestural score for words such as *bet* vs. *bed*, the perceptual importance of vowel length can be recognized along with the useful correspondence of the voicing difference to the orthographic representation.

In Bybee (1985, 1988b), I have shown how a network model with lexical connections among related words yields an internal morphological analysis and also allows for the recognition of submorphemic units, such as phonaesthemes and formatives such as *-ceive*. The mechanisms that lead to the establishment of lexical connections (relations of similarity or identity) and the formation of schemas is the same whether we are dealing with phonology or morphology. The substance involved—phonetic material or grammatical/semantic material—determines the differences in the emergent structures.

10. Location of "Universals" Compared to Optimality Theory and Natural Phonology

An important way that linguistic theories can differ from one another is in the status assigned to language universals. Much theorizing over the last few decades has followed Chomsky's very interesting suggestion that similarities across

languages emanate from the innate language faculty possessed by all human beings and that these innate principles can be specified in Universal Grammar. The principles residing in Universal Grammar operate as part of the language-specific grammar, but these principles are inborn and do not have to be acquired.

In early generative phonology most of these principles were structural and involved levels and ordering of rules, tiers and organizational features. Stampe's Natural Phonology proposed that substantive processes were also innate and showed up as applying in child language before the suppression of the ones that were not relevant for the language the child was acquiring (Stampe 1973). Optimality Theory also proposes that substantive universal constraints interact in grammars to evaluate the optimality of language-specific forms (Prince and Smolensky 1993).

My own view of the role of universals is quite different. While I would not deny that there are many innate capabilities that affect grammar, and that some of these may even operate in the day-to-day processing of language, the view that cross-linguistic similarities are accounted for once they are listed as part of UG misses the subtle and complex interaction of innate universals with language-specific systems.

Let us try to delve into the problem of what exactly a universal constraint is. As our example, let us take one that is quite uncontroversial: what Prince and Smolensky (1993) call "The Jakobson Typology" or the universal preference for CV syllables. The typology can be broken down into two parts which can be stated as preferences (or as absolutes, as in Optimality Theory): there is a preference for syllables to have onsets; there is a preference for syllables not to have codas. These universals are based on our knowledge about possible syllables in the languages of the world: all languages allow syllables with onsets but some languages do not allow V-initial syllables; all languages allow open syllables, but some languages do not allow closed syllables. The preferences, or constraints, are just summary statements about what can be observed in the languages of the world.

So far we are safe because we have done nothing more than state an observation about the languages of the world. It is the next step, actually a huge leap, that I find suspicious: taking a description of cross-linguistic facts and elevating it to an innate principle of Universal Grammar. This step resembles the problem mentioned in the Grocery Store Analogy: observing that grocery store check-out lines are usually the same length is one thing, making this a principle inherent in the structure of all grocery stores is quite another. In jumping from observation to universal principle, we have left out at least one step: we have not

asked how grocery store lines get to be the same length, and we have not asked how syllables get onsets and lose codas.[10]

A theory of universals must include a diachronic dimension (Bybee 1988a). For instance, Vennemann's (1988) Preference Laws for Syllable Structure are statements about preferences for change that in turn create synchronic states that follow certain implicational universals. The preference for CV syllables is described in two Laws, the Head Law and the Coda Law (Vennemann 1988: 13–14, 21):

> *Head Law*: A syllable head is more preferred: (a) the closer the number of speech sounds in the head is to one, (b) the greater the Consonantal Strength value of its onset, and (c) the more sharply the Consonantal Strength drops from the onset toward the Consonantal Strength of the following syllable nucleus.
>
> *Coda Law*: A syllable coda is more preferred: (a) the smaller the number of speech sounds in the coda, (b) the less the Consonantal Strength of its offset, and (c) the more sharply the Consonantal Strength drops from the offset toward the Consonantal Strength of the preceding syllable nucleus.

These Laws predict the directionality of phonological change and thus predict possible synchronic states. They are hypotheses or generalizations about syllable structure made by a linguist-observer. It is not proposed that these statements be incorporated into the grammar — they are not part of Universal Grammar, nor are they necessarily found in any language-specific grammar. They are not in themselves the real universals. The real universals are in the mechanisms that underlie the processes of change. The source of these universals is the architecture of the production and perception systems through which our cognitive structures are constantly funneled.

The last word on why syllables have an asymmetrical organization is not yet in, but it has long been suspected that it has something to do with the concentration of greater energy at the beginnings rather than at the ends of syllables. Browman and Goldstein (1992, 1995) find that the gestures of syllable-initial consonants tend to occur simultaneously while those of syllable-final consonants tend to be sequenced such that the gestures with the least constriction start before those with the most constriction. In addition, syllable-final gestures are also often reduced in magnitude compared to syllable-initial gestures. Thus the last gesture of a syllable may be overlapped and masked perceptually by the onset of the next syllable, leading to deletion. So the real explanation, the real universals, are way downstream from the lexicon and grammar: they are in the features of the

production and processing systems that mold and remold phonological material. It is highly probable that multiple physical and neural principles work together to create the typological effect described as a preference for CV syllables. There is not just one general constraint lodged in the innate cognitive apparatus and born into every human child. Instead, in this case a cluster of concrete, flesh-and blood facts about the vocal apparatus and its operation are responsible for affecting the phonological shape of syllables.

The moral to this story is that cross-linguistic generalizations are observations that we can make about language but they are not necessarily the same as the innate cognitive system that is used for language. Some universals come from phonetic factors, others arise because of the external context in which language is used, others from cognitive or perceptual factors that are independent of language. Only if language is viewed in the more general context of real usage by real language users will it become clear how to describe and explain cross-linguistic patterns.

11. Conventionalization in Phonology

Mark Durie (personal communication) has pointed out that the Grocery Store Analogy needs another chapter added to it, since it is possible for aspects of grocery-store behavior to become conventionalized, just as it is possible for certain aspects of pragmatically-determined linguistic behavior to become conventionalized. One way this could happen is for the grocery store manager to decide that all check-out lines had to be the same length, and shoppers could be directed to the shortest line. This situation corresponds to a language-level teleology and is not parallel to the way conventionalization takes place in language at all. The other scenario, which is parallel to linguistic conventionalization, is that the shoppers grow so accustomed to going to the shortest line that they forget what the original motivation was, and always go to the shortest line even if there is, for instance, an express line, which is longer, but in the end would be faster. This shopper-level conventionalization is parallel to the speaker-level conventionalization that takes place in language.

In phonology we can understand the presence of a phonological process in a language as the conventionalization of a natural phonetic tendency. For example, given a natural tendency to anticipate gestures, a velic opening will have a tendency to occur before the closure associated with a syllable-final nasal consonant. This anticipation may be present in all languages. However, in some

languages the anticipation may become conventionalized as part of the gestural timing repertoire that constitutes the phonetic character of the language. Then vowels before tautosyllabic nasal consonants will always be nasalized to some degree. Further anticipation and conventionalization of its effects may increase the extent to which the vowel is nasalized and eventually the nasalization on the vowel will become conventionalized as an inherent feature of the vowel, which then will remain even if the conditioning consonant itself deletes.

Thus we need the notion of conventionalization to explain why certain phonological processes are common but not universal, and what the relation is between phonetic naturalness or motivation and phonological regularities in a language. Other models are not so successful in representing these relations. If phonetic processes or constraints are innately given, then we have to explain why they are not present in all languages by saying that children must learn to suppress some processes or to order them with respect to one another. Furthermore, if they are innate, then the relation between the processes and their phonetic motivation is more distant and must make reference to the evolution of the species. If, as proposed here, phonological processes are conventionalized out of phonetic tendencies, then the relation is more direct.

12. Conclusion and Summary

In closing I will review some of the important points about usage-based functionalism as it applies to phonology. The first and most basic point is that language use shapes the grammar and lexicon. Frequent use of words and phrases leads to automatization and phonological reduction as well as to entrenchment or lexical strength in individual items, which makes them resistant to change to conform to more general patterns. High levels of applicability of a pattern to different lexical items lead to productivity of patterns.

The second point is that linguistic capabilities are not presumed to be different in structure from other cognitive capabilities. Linguistic units are stored like other percepts that come from our experience — detail is not factored out and the association between sound, meaning and context is direct, not filtered through intervening layers of structure. Stored linguistic units are categorized and associations are formed among them on the basis of their phonetic and semantic properties. Thus there is no separation of lexicon and grammar, lexicon and phonology.

In the recent era of intense development of synchronic phonology, it has not been the practice to study the units or patterns of phonology in the context of actual language use. Even phoneticians, who are closer to the data, tend to study utterances such as "Say mub twice" and "My pop puddles." The notion of language use as shaping phonological patterns has stayed in the background, invoked occasionally to explain sound change, but never viewed as an inherent part of the linguistic system. I suggest that the study of phonetic variation and detail, especially as they interact with lexicon and grammar, is likely to yield new and fascinating insights into a very subtle and complex pattern of human behavior.

Notes

1. The term "Usage-Based" comes from Langacker (1988). It is particularly felicitous because it unites the cognitive and functional perspectives. What Noonan (this volume) says about West Coast Functionalism applies to what I am calling Usage-Based Functionalism.
2. Actually not all "West Coast Functionalism" has been completely free of language-level teleology, but when pressed, I think proponents of this theory would all say that it is the speakers who have a purpose in mind, not the language.
3. Though I will not discuss it here, it is reasonable to propose also that the study of phonetic motivation for processes could also benefit from a diachronic perspective.
4. The vowel laxing is also used in the verbal paradigm (see Alonso, Vicente and Canellada de Zamora 1950).
5. The table presents percentages for convenience. The chi-squared value was not computed on the percentages.
6. As D'Introno and Sosa (1986) have pointed out, it is not just frequency that determines the rate of lexical diffusion of a phonological change, but familiarity, the frequency of a word in the social context where reduction and other changes take place.
7. In Francis and Kučera (1982), the following frequencies per million are listed: *despair* 20, *desperate* 26; *devout* 4, *devotion* 21.
8. The responses and the number of times they occurred were: *rintió* 12, *rentió* 6, *rentó*, 3, *rentuó* 1; *surnió* 1, *sornió* 20, *sornó* 1.
9. Francis and Kučera (1982) list the frequency of these words as follows: *damn* (57), *damnation* (3), *condemn* (30), *condemnation* (7), *hymn* (15), *hymnal* (0).
10. In fact, in Optimality Theory it is the principle itself that is responsible for giving syllables onsets and depriving them of codas, a situation that to me seems quite circular.

References

Alonso, D., A. Z. Vicente & M. J. Canellada de Zamora. 1950. "Vocales Andaluzas". *Nueva Revista de Filología Hispánica* 4:209–230.

Anderson, John R. 1993. *Rules of the Mind*. Hillsdale, N.J.: Lawrence Erlbaum.

Aske, Jon. 1990. "Disembodied Rules vs. Patterns in the Lexicon: Testing the Psychological Reality of Spanish Stress Rules". *Berkeley Linguistic Society* 16:30–45.

Boyland, Joyce Tang. 1996. *Morphosyntactic Change in Progress: A Psycholinguistic Approach*. Dissertation: University of California at Berkeley.

Browman, Catherine P. & Louis M. Goldstein. 1991. "Gestural Structures: Distinctiveness, Phonological Processes, and Historical Change". In I. Mattingly & M. Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, N.J.: Lawrence Erlbaum, 313–338.

Browman, Catherine P. & Louis M. Goldstein. 1992. "Articulatory Phonology: An Overview". *Phonetica* 49:155–180.

Browman, Catherine P. & Louis M. Goldstein. 1995. "Gestural Syllable Position Effects in American English". In F. Bell-Berti & L. J. Raphael (eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. Woodbury NY: American Institute of Physics, 19–34.

Bybee, Joan L., Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bybee, Joan L. & Joanne Scheibman, Ms. 1997. "The Effect of Usage on Degrees of Constituency: The Reduction of *Don't* in English".

Bybee, Joan L. & Elly Pardo. 1981. "On Lexical and Morphological Conditioning of Alternations: A Nonce-Probe Experiment with Spanish Verbs". *Linguistics* 19:937–968.

Bybee, Joan L. & Carol L. Moder. 1983. "Morphological Classes as Natural Categories". *Language* 59:251–270.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Philadelphia: Benjamins.

Bybee, Joan L. 1988a. "The Diachronic Dimension in Explanation". In J. Hawkins (ed.), *Explaining Language Universals*. Oxford: Basil Blackwell, 350–379.

Bybee, Joan L. 1988b. "Morphology as Lexical Organization". In M. Hammond & M. Noonan (eds.), *Theoretical Morphology*. San Diego: Academic Press, 119–141.

Bybee, Joan. 1995. "Regular Morphology and the Lexicon". *Language and Cognitive Processes* 10 (5):425–455.

Bybee, Joan. 1997. "The Phonology of the Lexicon: Evidence from Lexical Diffusion". In M. Barlow & S. Kemmer (eds.), *Usage-Based Models of Language*. Stanford: CSLI.

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

D'Introno, Francisco & Juan Manuel Sosa. 1986. "Elisión de la /d/ en el Español de Caracas: Aspectos Sociolinguisticos e Implicaciones Teóricas". In Rafael A. Núñez Cedeño, Iraset Páez Urdaneta y Jorge Guitart (eds.), *Estudios Sobre la Fonología del Español del Caribe*. Ediciones La Casa de Bello, 135–163.

Daugherty, K. G. & M. S. Seidenberg. 1994. "Beyond Rules and Exceptions: A Connectionist Modeling Approach to Inflectional Morphology". In S. D. Lima, R. L. Corrigan & G. K. Iverson (eds.), *The Reality of Linguistic Rules*. Amsterdam: John Benjamins, 353–388.

Dressler, Wolfgang U., et al. 1987. *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins.

Dressler, Wolfgang U. 1990. "The Cognitive Perspective of 'Naturalist' Linguistic Models". *Cognitive Linguistics* 1:75–98.

Durie, Mark. 1997. *Language, Function and Time*. Oxford: Oxford University Press.

Durie, Mark. This volume. "The Temporal Mediation of Structure and Function".

Fidelholtz, James. 1975. "Word Frequency and Vowel Reduction in English". *Chicago Linguistic Society* 11:200–213.

Francis, W. Nelson & Henry Kučera. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Goldberg, Adele E. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Haiman, John. 1994. "Ritualization and the Development of Language. In W. Pagliuca (ed.), *Perspectives on Grammaticalization*. Amsterdam: John Benjamins, 3–28.

Hall, Tracy Alan. 1989. "Lexical Phonology and the Distribution of German [?] and [x]". *Phonology* 6:1–17.

Harris, James W. 1969. *Spanish Phonology*. Cambridge, Mass.: MIT Press.
Hooper, Joan B. 1976a. *Introduction to Natural Generative Phonology*. New York: Academic Press.
Hooper, Joan B. 1976b. "Word Frequency in Lexical Diffusion and the Source of Morphophonological Change". In W. Christie (ed.), *Current Progress in Historical Linguistics*. Amsterdam: North Holland, 96–105.
Johnson, Theodore. 1983. *Phonological Free Variation, Word Frequency and Lexical Diffusion*. Seattle: University of Washington Doctoral Dissertation.
Kaisse, Ellen. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. San Diego: Academic Press.
Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
Kiparsky, Paul. 1982a. *Lexical phonology and morphology*. Linguistics in the morningcalm, ed. by I. S. Yang, 3–91. Seoul: Hanshin.
Klausenberger, Jürgen. 1979. *Morphologization: Studies in Latin and Romance Morphology* [*Linguistische Arbeiten* 71]. Tübingen: Max Niemeyer Verlag.
Klausenberger, Jürgen. 1988. "Toward Explanation in Morphology?". *Romance Philology* 41:409–419.
Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Basil Blackwell.
Langacker, Ronald. 1988. "A Usage-Based Model". In B. Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 127–161.
Langacker, Ronald. 1995. "A Dynamic Usage-Based Model". Paper presented at the Symposium on Usage-based Models, Rice University, March, 1995.
Langacker, Ronald. 1987. *Foundations of Cognitive Grammar, Vol. 1. Theoretical Prerequisites*. Stanford: Stanford University Press.
Leopold, W. F. 1948. "German ch". *Language* 24:179–180.
Leslau, Wolf. 1969. "Frequency as a Determinant of Linguistic Change in the Ethiopian Languages". *Word* 25:180–89.
Lindblom, Björn, Peter MacNeilage & Michael Studdert-Kennedy. 1983. "Self-Organizing Processes and the Explanation of Phonological Universals". *Linguistics* 21:181–203.
Lindblom, Björn. 1992. "Phonological Units as Adaptive Emergents of Lexical Development". In Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press, 131–163.
Lobben, M. 1991. *Pluralization of Hausa Nouns, Viewed from Psycholinguistic Experiments and Child Language Data*. M. Phil. Thesis, University of Oslo.

Lukatela, G. et al. 1980. "Representation of Inflected Nouns in the Internal Lexicon". *Memory and Cognition* 8:415–423.
Macfarland, Talke & Janet Pierrehumbert. 1991. "On ich-Laut, ach-Laut, and Structure Preservation". *Phonology* 8:171–180.
Miller, Joanne. 1994. "On the Internal Structure of Phonetic Categories: A Progress Report". *Cognition* 50:271–285.
Moder, Carol Lynn. 1992. *Productivity and Categorization in Morphological Classes*. SUNY at Buffalo Dissertation.
Moonwomon, Birch. 1992. "The Mechanism of Lexical Diffusion". Paper presented at the Annual Meeting of the Linguistic Society of America, January 1992, Philadelphia.
Moulton, William. 1947. "Juncture in Modern Standard German". *Language* 23:212–216.
Noonan, Michael. This volume. "Non-Structuralist Syntax".
Nosofsky, R. M. 1988. "Similarity, Frequency, and Category Representations". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:54–65.
Ohala, John. 1992. "The Segment: Primitive or Derived?". In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 166–189.
Ohala, John J. & Manjari Ohala. 1994. "Speech Perception and Lexical Representation: The Role of Vowel Nasalization in Hindi and English". In Bruce Connell & Amalia Arvanti (eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge: Cambridge University Press, 41–60.
Phillips, Betty S. 1984. "Word Frequency and the Actuation of Sound Change". *Language* 60:320–342.
Pinker, Steven. 1991. "Rules of Language". *Science* 253:530–535.
Pisoni, David B, Howard C. Nusbaum, Paul A. Luce & Louisa M. Slowiaczek. 1985. "Speech Perception, Word Recognition and the Structure of the Lexicon". *Speech Communication* 4:75–95.
Prince, Alan & Paul Smolensky. 1993. "Optimality Theory". Technical Reports of the Rutgers University Center for Cognitive Science, Vol. 2.
Scheibman, Joanne. To appear. "*I Dunno But* ... A Usage-Based Account of the Phonological Reduction of 'Don't' in Conversation".
Stampe, David. 1973. *A Dissertation on Natural Phonology*. University of Chicago doctoral dissertation.

Studdert-Kennedy, Michael. 1983. "On Learning to Speak". *Human Neurobiology* 2:191– 195.

Studdert-Kennedy, Michael. 1987. "The Phoneme as a Perceptuomotor Structure". In A. Allport, D. MacKay, W. Prinz & E. Scheerer (eds.), *Language, Perception, and Production*. New York: Academic Press, 67–84.

Tranel, Bernard. 1981. *Concreteness in Generative Phonology: Evidence from French*. Berkeley and Los Angeles: University of California Press.

Vennemann, Theo. 1972. "Rule inversion". *Lingua* 29:209–42.

Vennemann, Theo. 1988. *Preference Laws for Syllable Structure*. Berlin: Mouton De Gruyter.

Wang, H. S. & Derwing, B. L. 1994. "Some Vowel Schemas in Three English Morphological Classes: Experimental Evidence". In M. Y. Chen & O. C. L. Tzeng (eds.), *In Honor of Professor William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Taipei: Pyramid Press, 561–575.