

A view of phonology from a cognitive and functional perspective¹

JOAN L. BYBEE

Abstract

While morphosyntax and semantics have been studied from a functional and cognitive perspective, much less emphasis has been placed on phonological phenomena in these frameworks. This paper proposes a rethinking of phonology, arguing that (i) the lexical representation of words have phonetic substance that is gradually changed by phonetic processes; (ii) the spread of these phonetic changes is at least partly accounted for by the way particular items are used in discourse; (iii) the study of exceptions, marginal phenomena, and subphonemic detail are important to the understanding of how phonological information is stored and processed; (iv) generalizations at the morphological and lexical level have radically different properties than generalizations at the phonetic level, with the former having a cognitive or semantic motivation, while the latter have a motor or physical motivation; and (v) that the best way to model the interaction of generalizations with the lexicon is not by separating rules from lists of items, but rather by conceiving of generalizations as patterns or schemas that emerge from the organization of stored lexical units.

1. Introduction

In the last two decades there has been an active interest in viewing morphosyntactic phenomena from a typological, functional, and cognitive perspective. These research perspectives have proved extremely productive in pushing us closer to an explanation for the grammatical structures that exist in the languages of the world. Research on universals, discourse, conversation, and cognition have stretched the boundaries of linguistics and insisted on the principle that linguistic structures are dependent upon semantic and cognitive substance and the uses to which languages is put. But no movement of comparable strength has emerged which applies this principle to phonological phenomena, even though the

relevance of the application is obvious: phonological structures must be dependent upon substance—phonetic substance, and they are surely molded by the uses to which they are put.

Phonological research in these same two decades has taken a number of interesting steps—the inventory of basic units has been expanded to include the syllable; the nature of the adherence of features to particular segments has been re-examined; suprasegmental phenomena have been seriously scrutinized; and the interaction of syntax with phonology has been studied closely. Throughout these developments the goals of phonological theory have remained static and the basic structuralist assumptions of phonology have gone largely unquestioned. The existence of abstract underlying representations is still assumed, as is the existence of rules which make changes to these representations; it is still assumed that underlying systems are regular and symmetrical and that the prime units in phonology are distinctive features.²

In the inaugural volume of this journal, Dressler argued that Natural Phonology, indeed all of naturalist research, is based on a version of functionalism, which he describes as follows (Dressler 1990: 76):

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

In this view language structures or phenomena are matched with a semiotic function which serves as the explanation for their existence.

The type of functionalism that has been pursued lately in American research into the relation of discourse to morphosyntax (e.g. in papers such as Hopper and Thompson 1984; Du Bois 1987) and in studies of grammaticization does not assume that grammar is created to serve certain functions, but rather that grammar is the conventionalization of frequently-used discourse patterns. In this view, grammar is an artefact of the communication process, and the properties of natural language are molded by language use.

The goal of the present paper is to adopt the latter type of functionalist view and outline some aspects of a program for the reconsideration of phonology in the light of that view. Of course, the total re-examination of phonology will be many-faceted and occupy numerous research lifetimes, but it is hoped that by sketching some points in this direction others will be stimulated to look at phonology afresh. I apologize in advance for the fact that this is not a typical research paper, in the sense that I do not report on original research, nor do I attempt to fully survey work already published in the areas I discuss. Rather I am attempting

to articulate a certain point of view about phonology that arises from the application of functionalist (in the sense outlined above) principles to phonology. This perspective gives rise to the following mandates:

— *Consider the substance of phonology rather than just the structure.* For syntax the substance is word meaning and grammatical meaning. For phonology the substance is largely phonetic, but generative phonology has also allowed the consideration of morphology and lexicon in the conditioning of “phonological” rules. It is our task to distinguish the type of substance involved in any particular phenomena, for different substance conditions different types of behavior.

— *Consider the uses to which phonological elements are put.* Traditionally phonological variants are only considered in their “type distribution”—the phonological environments in which they occur—much as syntax was only studied in the rarefied environment of made-up sentences. Very little study has been devoted to the distribution of phonological elements in texts. I will argue below that the text frequency of segments affects their phonetic shape and evolution.

— *Consider subphonemic detail and variation conditioned lexically, morphologically and socially.* Generative phonology, like its predecessor, phonemic theory, chose to ignore “low-level phonetic detail”. Like the detail of actual language use that has enriched functionalist syntactic theory, the study of detail in phonology will reveal important facts that bear on our understanding of how language is really processed and what structures have empirical validity.

— *Attend to exceptions and marginal cases,* for they can be valuable sources of information about the nature of processing and representation. As I will argue below, marginal “phonemes” are particularly interesting in their consequences for phonological theory.

— *Reconsider what Langacker (1987) calls the “rule-list fallacy”* (see also Bybee 1988). Our thinking and analyses need not be restricted to only two options—either an element occurs in a list or it is generated by rule. I propose below that lexical elements (words or phrases) consist of actual phonetic content that is modified as these elements are used. While phonetic “rules” may exist as articulatory patterns for the realization of words, generalizations at other levels may be better thought of as emergent generalizations over lexical representations.

2. Properties of different alternation types

The starting point for phonological alternations are the natural coarticulation and reduction phenomena that occur when language is used. Such effects quickly become involved in the expression of meaning, since the

major function of language is the organization and transmission of meaning. In particular, phonetic alternations become associated with particular grammatical morphemes and often also with particular lexemes. It is important to distinguish the various stages of evolution of phonological alternations, because the nature of the generalizations differ according to whether the process is purely phonetic, morphological, or lexical.

Phonetic alternations are motivated by motor production and produce phonetic alternates that are minimally different from one another (e.g. American English [t] and [d] vs. the flap). Generalizations at this stage are storable in purely local phonetic terms and will here be referred to as phonetic processes. Given an appropriate notation (such as that found in Browman and Goldstein 1986), it will probably be possible one day to define substantively the maximal phonetic difference that can be phonetically conditioned, a proposal also made in *Natural Phonology* (see Dressler 1985: 64–65). That is, given a productive phonetic process, once the phonetic change effected by this process has progressed beyond a certain limit, the results of the process can no longer be phonetically conditioned (or allophonic) but are rather reanalyzed as belonging to morphological categories or lexical entries. To use an example that will be taken up again later, the palatal variant of German /x/, [ç], which is in most environments a phonetically predictable and natural variant, I would suggest has now moved too far from the velar to be phonetically conditioned. In Section 5 I will argue that it is in fact marginally phonemic. The difficulty with testing this claim is that reanalysis is covert and the surface evidence for this reanalysis may not appear until long after the change has occurred. Another difficulty in testing this claim is that it requires making much finer phonetic distinctions than phonologists have been accustomed to making.

An obvious objection to the claim that there are substantive differences between phonetic variants and variants that have been morphologized or entered in the lexicon is the apparent fact that voiced and voiceless alternations, for instance, can be of either type. Thus English has morphological and lexically conditioned alternations in words just as *house*, *houses*, *wife*, *wives*, and German and other languages appear to have the same voiced and voiceless alternation due to syllable-final devoicing. It has been shown, however, that the voiceless variant derived by this productive process in German has not lost all the properties of a voiced stop; it is not identical to the phonemic voiceless stop either in its own properties or in its affect on surrounding articulations (Port and O'Dell 1985). Given better phonetic description and a more constrained view of productivity, we will probably find that truly productive phonetic processes, which are minor adjustments in articulation, make very small

changes compared to those described by non-phonetic (lexical or morphological) generalizations.

To the extent that morphophonemic patterns or lexical generalizations are viable and productive, their patterning is quite different from that of phonetic processes. Phonetic processes occur in strictly phonetic environments and can be viewed as the reduction or overlapping of articulatory gestures; their properties will relate to the general properties of motor gestures and the particular features of the articulatory system (Browman and Goldstein 1986, 1990; Pagliuca and Mowrey 1987a).

In contrast to phonetic processes, morphologically-conditioned alternations tend to diagram the semantic distinctions made in the morphology; they undergo diachronic changes that show that the parameters relevant for morphology are quite different than for phonology (Bybee 1985). For instance, an alternation that is morphologized may take on different properties in nouns than in verbs. Spanish stress for example is morphologized for verbs and has undergone some changes motivated by verbal categories that are different from possible changes in nouns (Hooper 1976a). A class of sounds that has undergone a phonetic change together may break apart when morphological and lexical factors intervene. For example, the alternation of /e/, /ye/ and /i/ in Spanish verbs is marginally productive, while the back vowel counterpart alternation of /o/, /we/ and /u/ under the same morphological conditions is totally unproductive (Bybee and Pardo 1981).

Not only does the domain of generalizations change in morphologization but the directionality of generalizations also changes. The basic-derived relation in morphology is based upon the semantically basic-derived relation: alternations are predicted from singular to plural, from present to past, from third person to other persons (Bybee and Brewer 1980; Bybee 1985), no matter what the original distribution of conditioning was (Vennemann 1972). As argued by Greenberg (1966), Mancaz (1980), Bybee and Brewer (1980), and Bybee (1985), these relations are based in large part on the way in which forms are used: the more frequently used forms are the ones that are taken to be more basic. The directionality of phonetic processes, in contrast, is from a gesture unaffected by surrounding gestures to one that is reduced or overlapping or deformed by contiguous gestures.

Lexical generalizations evince yet another pattern. Lexical classes of verbs or nouns with special morphological characteristics or stress patterns (in languages in which stress is partially lexicalized) resemble natural categories. That is, they have the shape of generalizations that human beings make about non-linguistic categories, which include the ability to make use of holistic features of words as well as local ones, to group

words in specific as well as general patterns and to impose a prototype structure on class members (Bybee and Moder 1983; Aske 1990; see also Section 6 below).

Finally, we can consider evidence available from the process of lexical diffusion—the process by which changes spread through the lexicon—as lexical diffusion provides evidence for the way that phonology and morphophonology interact with the lexicon. Here we see again a difference between phonetic and non-phonetic processes: phonetic changes spread gradually through the lexicon and tend to be more advanced in more frequent words. Non-phonetic alternations have a tendency to be lost (leveled) and this loss occurs earliest in less frequent items. More frequent items tend to be more conservative in non-phonetic alternations but more innovative in phonetic ones (Hooper 1976b). Phonetic alternations are both phonetically and lexically gradual, but they do not usually have exceptions—eventually all words will fall prey to phonetic change. Non-phonetic alternations not only can have exceptions—they usually *are* the exceptions.

3. *Phonetic substance*

The phonetic material that forms the substance of phonology can be viewed in two new ways: first, rather than abstracting away and regularizing in the name of establishing a small and coherent set of distinctive features, the actual phonetic shape of linguistic units (whether at the level of the segment or at higher levels) should constitute the object of study; second, words can be regarded as consisting of real phonetic substance, in the lexicon as well as in production and perception, and this phonetic substance can be regarded as undergoing permanent change as the result of coarticulation processes.

3.1. *Distinctive features*

As their name implies, distinctive features are not very useful in understanding the nature of phonetically-conditioned processes. Since they were developed to describe contrast they often lack the level of detail necessary to describe (and thus understand) details of coarticulation.

In addition, like other componential features from structuralist theories, their constancy across manifestations in different combinations is a fiction. Just as “plural” does not mean the same thing in first person as it does in third person (since an inclusive/exclusive distinction is possible in first plural but not in third), so [continuant] does not mean the same thing for labials as it does for alveolars. There is some commonality, of course, in that a total obstruction of the vocal tract is attained in noncon-

tinuants, but the gestures required to accomplish this closure are completely different depending upon the point of articulation (Browman and Goldstein 1986).

This problem is more marked for the really arbitrary features such as [anterior], but even [continuant] should come under scrutiny given its different realizations in different environments. The evidence for such a feature is its ability to define a “natural class”; the received position is that classes of stops (either voiced or voiceless) tend to behave in a uniform manner across a wide variety of languages. Of course, exceptions are well-known: [p] tends to spirantize and delete while [t] and [k] remain intact (e.g. in Japanese, in many Altaic dialects, in Eskimo-Aleut [Pagliuca and Mowrey 1987b]). Or consider the changes undergone by Proto-Bantu voiceless stops (Tucker and Bryan 1957; Pagliuca and Mowrey 1987b):

Proto-Bantu	*p	*t	*c	*k
Ganda	w	t	s	k
Gikuyu group	h	t	ð	k
Pokomo	ϕ	h	(d)s	k
Luhya dialects	h	(r)t	s	x

In the face of these exceptions, we might want to ask whether there is strong evidence for the feature [continuant] in the form of “natural classes” of stop consonants. Pagliuca and Mowrey (1987b) investigated this question using a sample of 39 unrelated languages. They examined all allophonic processes that affect voiced and voiceless stops in these languages. They counted as “regular” a process that affects the whole series, that is, all voiceless stops or all voiced stops, in the same way. Their counts revealed that fewer than 50% of the allophonic processes affecting a stop series were regular in their sense. Only 49% were regular for voiceless stops and 42% were regular for voiced stops. This investigation suggests that the evidence for natural classes is considerably weaker than has been supposed, and shows that it is important to submit our theoretical constructs to empirical testing where possible. More cross-linguistic research may reveal that some features define classes better than others. In the meantime, Pagliuca and Mowrey suggest giving up the “building block” model of phonology, as their data casts doubt on the reality of features and segments. Rather, they argue that “phone-

sized" units are heavily affected by context and that a model of holistic processing for words and phrases is more realistic, a point to which we now turn.

3.2. *The phonetic substance of lexical items*

On the phonetic side, words or larger lexical units are auditory percepts associated with motor command patterns. What we know about other stored perceptual units suggests that detail is not factored out, but rather preserved to a large extent. The same must be the case with the phonetic representation of words. As Langacker (1987: 391–392) points out, if the phonetic regularities in words are factored out, the residue would not be "recognizable as a coherent entity plausibly attributed to cognitive autonomy." In other words, if we suppose that the perceptual representation of linguistic units, such as words, are stored in a way that is analogous to the way we store other units of perception, such as the visual representation of natural objects, we would not suppose that redundancies are removed for storage. What this would mean, for example, for the visual representation of a human being is that the arms and legs, eyes, ears, nose, etc. could all be removed since they are predictable (with some exceptions) from the feature [+human]. If we recognize stored linguistic units as objects with cognitive autonomy, then they require a full and rich representation, with recognized regularities in their properties represented as schemas at different levels.³

Moreover, many empirical linguistic facts are not explainable unless words are viewed as consisting of actual phonetic substance. Under this proposal, phonetic effects of coarticulation and reduction represent changes that are made to the actual tokens, which also affect the stored representations of words. Each new perceptual instance of an item—whether the input is from another person or from self-monitoring—is superimposed on existing representations and will gradually change the representation. We will consider three examples here that support this claim.

Consider some English words first discussed by Zwicky (1972) and later in Hooper (1976b, 1978). These are words that end in two unstressed syllables, the penultimate containing schwa followed by [r] or [l].

- In many of these words the schwa has been deleted: *every, camera*.
- In other words, however, the schwa is variably present, or at least a syllabic [r] or [l] are found to alternate with the absence of syllabicity: *memory, celery, summary, family*.
- At the far end of the continuum are words which do not have schwa deletion at all; a full syllable is always present: *mammary, artillery*.

In Hooper (1976b) I showed that the presence, variability or absence of the schwa correlated with the frequency of the word. The words of highest frequency have undergone schwa deletion completely, while the lowest frequency words do not allow the reduction at all. The role of word frequency will be discussed again in Section 4 below.

What this continuum shows is that the degree to which the schwa is present is associated with a particular word or lexical unit; it is not phonetically or morphologically predictable, nor is it determinable from social factors alone. The difference in the pronunciation of *memory* and *mammary* is something that must be learned about these individual words. It is, however, something that follows directly from the input and is learned as the word is learned. If *mammary* is always heard with three syllables but *memory* is heard sometimes with two and a half and sometimes with two, that is the way these words are stored. Just as detail about the meaning and co-occurrence patterns of words is acquired and stored with the word, details about its phonetic representation are also acquired and stored.

It is most reasonable to suppose that words with this phonetic environment form a continuum as each one is gradually succumbing to greater and greater reduction of this unstressed syllable. But even if we recognized three discrete groups of words—those with full deletion, those with syllabic [r] or [l] and those with schwa plus [r] or [l]—this case would still present a problem for structural theories of phonology, since syllabic [r] and [l] and schwa plus [r] and [l] do not contrast elsewhere. At the very least, one could argue that this example suggests that some rather fine phonetic distinctions can be associated with particular lexical items, which would naturally follow from the postulate that lexical items consist of real phonetic substance.

A second example of lexical diffusion of a gradual phonetic change illustrates another way that change illustrates another way that change in the phonetic shape of words can affect surface phonology. The loss of syllable-final [s], especially in the dialects of Latin America, has been well-studied. The /s/ tends to lose its lingual articulation, remaining only as voicelessness (called aspiration and represented as [h]), and it eventually deletes in syllable-final position, that is, before a consonant. This occurs regularly even across word boundaries, but since Spanish tends to resyllabify across word boundaries, a word-final /s/ before a vowel in the next word would be syllable-initial and would not delete. The following illustrate this:

- word-internally before a C: *felizmente* [felihmente] 'happily'

— word-finally before a C: *o se traen animales finos* [animaleh finos] 'or fine animals are brought'

— word-finally before a V: *no vas a encontrar* [... vasa ...] 'you aren't going to find'

In strings of words, a word-final /s/ will sometimes appear before a vowel and sometimes before a consonant. The pre-consonantal environment is two to three times more frequent than the pre-vocalic environment. Thus the word-final /s/ is often in the environment for reduction, and our hypothesis would predict that this will have a permanent effect on that /s/. The data support this hypothesis: as /s/-reduction proceeds, word-final /s/ is eventually affected, even when it occurs before a vowel in the next word. This is not a phonetically-conditioned generalization of the process, because /s/ before a vowel inside a word is never aspirated; the aspiration of /s/ before a vowel takes place *only* word-finally. The fact that the description of the situation requires that we mention word-final position in itself suggests that the reduction of this /s/ has to do with the fact that it appears in particular words in a particular position. It might be added that word-final /s/-reduction proceeds at different rates in different words (Terrell 1975, 1977; Hooper 1981).

Another very productive, ongoing reduction process in Spanish is the spirantization of voiced stops following a vowel. The interaction of spirantization with /s/-reduction provides a third example showing that actual phonetic substance in lexical items is substantively affected by phonetic processes. If a syllable-final /s/ before a voiced stop were to delete entirely, the voiced stop would be preceded by a vowel and the right conditions for the application of spirantization would arise. Amastae (1989) reports on this situation in Honduran Spanish. His data show that the voiced stop is no more likely to spirantize if the /s/ is deleted than if it is fully there. There is no question about the productivity of spirantization, so one would expect that the newly post-vocalic stop would instantly spirantize. However, if stops in these words had been protected from reduction by the preceding /s/, and if reduction is a gradual process that affects lexical representations, then these newly exposed stops will take a while to spirantize.

I am claiming, then, that lexical representations are not abstract, unchanging entities that processes apply to but that in essence remain unchanged; rather, they consist of substance that is gradually but permanently modified in language use. The old metaphor for this in non-Neogrammarian historical linguistics was that words are like coins that are worn down as they are passed from hand to hand. The slogan that went along with this theory was "Every word has its own history".

This is not to deny the reality or systematicity of coarticulation patterns. Each language has associated with it a set of motor patterns or habits that strongly govern the phonetic output, referred to earlier as phonetic processes. My claim is simply that these same patterns have a permanent effect on the words that are filtered through them. Menéndez-Pidal (1950: 531) describes the interaction of phonetic patterns with lexical items using the analogy of leaves floating on a flowing river:

Cada palabra que en fonética parezca discordante de sus análogas, puede estar sometida a una tendencia general que la impulsa en unión con las otras. Todas son llevadas por la misma corriente, como multitud de hojas caídas en un río; cada hoja sigue su curso especial, tropieza acaso con obstáculos que la desvían, la retrasan o la detienen, pero todas están sometidas a la misma fuerza, ora la arrastre, ora solamente las empuje, y sería ceguera empeñarse en observar el curso de cada una sin darse cuenta de la corriente que las domina a todas.

[Each word whose phonetic form seems discordant with its analogues yields to a general tendency that drives it to a union with the others. All are carried by the same current, like a multitude of leaves fallen in a river; each leaf follows its own special course, running into obstacles that deflect it, delay it or detain it, but all yield to the same force, now the force drags it, now it pushes it, and it would be blindness to insist on observing the course of each one without attending to the current that dominates them all.]

The final argument (and perhaps the strongest one) for the hypothesis that words consist of phonetic substance that is permanently modified by phonetic processes is the fact that when phonetic processes become unproductive their effects are never undone. In English, the voicing of intervocalic fricatives has not been productive for a long time, as we can see by all the examples of French borrowings with voiceless intervocalic fricatives, e.g. *classes*, etc. Yet the fricatives affected by this process show no tendency to devoice—*of* is steadfastly maintaining its [v]. The only places we see a tendency for the old voiceless fricative to be restored are in cases of alternation, where the unmarked category has the voiceless fricative, as in *house*, *houses*, where the latter is sometimes heard among younger people with an [s] rather than a [z]. This restoration is due to morphological change, not phonetic change.

4. The role of frequency

The tendency in structural phonology, as in all applications of structuralism, is to concentrate on structural patterns and ignore patterns of actual use. In phonology, as in morphosyntax, this means neglecting an important source of explanation for distribution patterns. The role of word frequency has been consistently neglected in phonology and even in

variation studies, yet there is evidence that frequency of use is an important factor in the spread of phonological processes.

We have already seen an example of word frequency as a determining factor in phonetic reduction in the example of schwa reduction and deletion in certain English words. Given the postulate of words having actual phonetic substance, what we can say about this phonetic reduction is that the words are actually eroded in use and that this erosion is reflected in their gradually changing lexical representations. Thus frequency is important, not just as a predicting variable but as an actual causal mechanism, since words that are more frequent will be used in positions of lesser stress and emphasis and will thus be more susceptible to reduction processes.

There is also a second way that word frequency can be invoked as a causal mechanism in phonetic change. Word frequency is also a factor in the lexical diffusion of vowel shift changes, and it is not clear if any or all of such changes are reductive in nature. For instance, Moonwomon (1992) studies the backing of /æ:/ and the fronting of /a:/ in San Francisco English and finds that the two most frequent words with these segments, *class* and *got*, manifest the change to the greatest extent. Even if these changes cannot be considered reductive (though of course they might very well be reductive), their frequency can still be a factor in causing their advancement in the change. As Moonwomon points out, if productive processes apply in real time as speakers speak, more frequent words will have a greater opportunity to undergo a process than less frequent words. If the application of a process can have a permanent effect on a word, as I am suggesting here, the application of these processes, which produce only small incremental changes, will eventually accumulate a more noticeable mutation in the more frequent words.

Frequency as an explanatory factor can be used to answer some questions that phonologists have found imponderable to date. For instance, in Spanish spirantization, the labial, dental and velar stops all spirantize, but the dental stop spirantizes more regularly and becomes weaker or closer to deletion than the labial or velar stops. There seems to be no phonetic explanation of this fact. However, an explanation can be found in language use. The dental stop is much more frequent in texts than either of the other stops. Navarro Tomás (1946) reports that in a count of 20,000 phonemes in running text, /b/ accounts for 2.54% of the total, /g/ for only 1.04%, and /d/ for 5.00%, making /d/ twice as frequent as /b/ and nearly five times as frequent as /g/. In fact, /d/ is the most frequently occurring stop in Spanish, and among obstruents in general it is second in frequency only to /s/.⁴

I have cited phoneme frequency here, but I do not believe that it is the frequency of the segment itself that is important. It is rather the fact

that the /d/ occurs in *words* that are more frequent. The overall greater reduction of /d/ is due to its occurrence in some very frequent words (such as *lado* 'side') and morphemes (such as the past participle, *-ado*).

One difficulty with various theories of consonant weakening is that while the directionality for manner of articulation (e.g. from stop to fricative, from voiceless to voiced) are regular across language, which of the various points of articulation lead the process and which lag behind differs across languages (Foley 1977). In order to make theories of weakening more predictive across languages, more factors need to be taken into account; these include language-specific phonetic factors, and may also include the distribution of consonants in frequent versus infrequent words.

The role of frequency in the reduction of grammaticizing constructions is just a special case of erosion proceeding at a faster pace in frequent than in infrequent items. Most (although probably not all) of the phonetic processes occurring in the contraction of grammatical elements have a counterpart in the much slower-moving "normal" phonetic processes of the language. Thus the deletion of /d/ from the Spanish second plural morpheme, Latin *-atis*, Old Spanish *-ades*, Modern Castilian *-ais*, predated the general weakening of /d/ by many centuries, and would have seemed aberrant at the time. However, exactly the same change is slowly plodding through the Spanish lexicon today.

The contraction of *will not* to *won't* appears anomalous unless one considers the fact that a postvocalic /l/ before a consonant tends to vocalize to a back glide in current English, as in *milk* [miwk]. Then considering the rounded initial glide and reduced vowel followed by a back glide, the emergence of [o] is not too surprising. The advanced compression and reduction of *won't* is then due to frequency. It is also frequency that allows this negative contraction to diverge so much from its related affirmative form. That is, *won't* is not just fused phonologically—it has become an independent semantic and grammatical unit as well. It is its high frequency that allows this lexical autonomy.

These examples of reduction and contraction of grammatical elements are well-known and accepted (though seldom studied in their own right). What I am claiming is that this differential reduction due to frequency is pervasive throughout the forms of a language, though in a more subtle and less dramatic form. The study of patterns of use, then, is likely to add a new dimension to the study of phonology, perhaps answering some of the long-standing questions persisting in the field.

5. Marginal phonemes and how they arise

If phonological processes have a permanent effect on lexical representations, phonetic elements that are not phonemic will occur in lexical

representations. This in turn will lead to the development of new "phonemes", and also to problematic cases where a phone appears to be predictable in many, perhaps most, contexts but appears not to be predictable on phonetic grounds in a few contexts. A good example of this phenomenon of "marginal phonemes" (which probably occur in all languages) is the palatal fricative of German. In almost all contexts the palatal is a predictable variant of the velar fricative, occurring after front vowels and the consonants /r/, /l/ and /n/. However, the consistent realization of the diminutive suffix *-chen* contains the palatal fricative, even when the preceding noun ends in a back vowel, as in *Tauchen* [ç] 'small rope' and *Pfauen* [ç] 'little peacock'. (These form minimal pairs with the verbs *tauchen* 'to dive' and *fauchen* 'to spit', which have the velar fricative.) Further, certain loan words contain the palatal fricative in initial position or after a back vowel: *China* [ç] and *Photochemie* [ç]. The traditional solution to this problem (Moulton 1947; Leopold 1948) is to predict the [ç] from a boundary (or juncture). This is tantamount to claiming that the [ç] is part of certain morphemes and words, which in turn amounts to lexical status for [ç].

The consistent realization of *-chen* with a palatal, even after a back vowel, arose because this suffix was represented lexically with a palatal, at first even though the palatal was predictable. Eventually the palatal is so associated with this morpheme that it occurs even outside its original phonetic environment.

Another example of a "marginal phoneme" is the schwa in English. The schwa is clearly a reduced vowel, but in many cases it is difficult to say what it is a reduced version of. Similarly, the flap of American English is clearly derivable from a /t/ or /d/ in words like *writer* and *riders*, but is always a flap in nonalternating environments, such as in the words *butter*, *ladder*, and *latter*. In these words, and even in words such as *writer*, the flap is lexical.

Given the argument concerning lexical diffusion and the representation of phonetic detail in the lexicon, one could wonder why languages have small numbers of contrastive units rather than having a number of phonetic representations equal in size to the number of words in the language. Why do languages have repeated patterns that can be described as phonemes and as syllable-structure patterns rather than a totally different set of phonetic parameters for each word? There are two reasons, one involving cognitive processing and one involving motor production.

The processing of words, like that of other percepts, involves a sorting and matching with other similar stored images. This means that phonetically similar elements of words will be associated with one another and common patterns will be organized into schemas. These schemas repre-

sent motor patterns that are used in actual production. As in any motor task that takes place with great speed and fluency, repeated patterns facilitate production. There is a tendency, then, to limit, through actual use, the number of distinct motor patterns necessary to produce speech in a given language.

6. The cognitive domain: lexical patterning

In the preceding I have been arguing for lexical storage of phonetic detail and variation where the associated processes have an articulatory *raison d'être*. In principle, these processes could be completely different in quality from "rules" or generalizations at higher levels of cognitive functioning. However, since so little is known about the psychological functioning at any of these levels, I will not attempt a comparison, but in this section I will cite research that reveals lexical patterning at a level above the phonetic. In particular, I will discuss Spanish stress and English verb classes as matters of lexical patterning rather than as generated by a "disembodied" or "symbolic" rule that is independent of the items to which it applies.

In Bybee (1988) I argued that generalizations that have lexical and grammatical conditioning are not separable from the lexical items that they affect. In fact, in many cases it is very difficult to get speakers to apply what have been described as "rules" to new items. My claim is that the generalizations described by these rules are not extracted and stored in a separate component, but are rather patterns or schemas that emerge from connections among stored units. Such a theory accommodates in a natural way the fact that patterns with lexical and grammatical conditioning range widely from the very specific to the very general and readily tolerate exceptions.

Aske (1990) argues from experimental data that Spanish stress is best treated as a lexical pattern rather than a "disembodied rule". Spanish stress is predictable in nouns and adjectives in over 95% of cases by the statement that words ending in vowels have penultimate stress, while words ending in consonants have final stress (note that the crucial reference to the end of the word means that the generalization requires grammatical information):

Penultimate		Ultimate	
<i>ventána</i>	'window'	<i>ladrón</i>	'thief'
<i>mercádo</i>	'market'	<i>amanecér</i>	'dawn'
<i>cálla</i>	'street'	<i>ciudad</i>	'city'

In generative phonology this statement is canonized as rule, and exceptions to it have to be handled lexically. In early generative phonology they are marked with an abstract phonological feature, in more concrete generative phonology with an arbitrary diacritic, and in metrical phonology certain vowels are marked as "extra-metrical" (meaning they are ignored for the purposes of stress assignment). However the exceptions are handled, they are all regarded as equally exceptional, having to be learned individually. Consequently, the rule analysis predicts that nonce or new words will be stressed according to the general rule.

Aske (1990) observes, however, that some more specific conditions may govern just how exceptional the exceptions are. While words ending in *-Vn* are overall stressed finally in the vast majority of cases, there are many familiar words ending in /en/ which have penultimate stress. In fact, among familiar words ending in /en/ more than half have penultimate stress (*orden* 'order', *imagen* 'image', *origen* 'origin', *abdomen* 'stomach'). Aske predicts that this small but strong group of words (with approximately 30 members) might make up a lexical pattern that could influence the way subjects stress new words. In an experiment in which he asked Spanish speakers to read nonce or rare words written with no accent indicated, he found that their responses to words ending in /n/ very accurately reflected the stress patterns of existing words ending in /n/. In particular, words ending in /en/ had penultimate stress in 43% of the cases and final stress in 55%. Words ending in other vowels plus /n/ received final stress in 96% of the cases.

Another similar case involves words ending in *-ico* or *-ica*, the vast majority of which have antepenultimate stress: *república* 'republic', *química* 'chemical', *fonética* 'phonetic'. These words would have to be marked as exceptional in the rule analysis, despite the fact that they form a subregularity among themselves. As predicted in the experiment, the subjects assigned the allegedly exceptional antepenultimate stress to words ending in *-ico* or *-ica* in 73% of the cases, showing a clear preference for antepenultimate stress over penultimate in this case.

These data show that noun and adjective stress in Spanish is not a matter of one general rule which has some exceptions; rather, the many alleged exceptions make up productive subpatterns. Note that Spanish stress evolved from a system in which one general rule was applicable to all words. The evolution and the current situation suggest that speakers prefer at least at some language stages to organize words into sub-groupings rather than to apply a general rule. This strongly suggests that Spanish stress is indicated in each lexical entry. Aske refers to this as "non-wasteful" listing because such listing makes possible the emergence of important subpatterns. What is characterized in generative phonology

as a rule is just the strongest and most general of these lexical patterns (Bybee 1988).

Very similar phenomena can be observed in lexical classes that are significant morphologically. While the English past tense in *-ed* is clearly productive, the various irregular verbs of English are not all equally irregular. The English verb class typified by *string/strung* is the only one that has continually gained new members since the Old English period. In nonce-probe tasks it is easy to get subjects to assign nonce verbs to this class, provided that these verbs have a phonological shape resembling the existing members of this class. Bybee and Moder (1983) have shown that membership in this class is based on similarity to a prototype, the strongest feature of which is that it ends in a velar nasal and a secondary feature of which is that it begins in a consonant cluster, preferable involving /s/. The evidence for the prototype or family resemblance structure of this class is the fact that, while the class originally included only members with final nasal clusters, it now includes members with final oral velars, such as *strike/struck* and *dig/dug*, showing that the prototype for the final consonants has shifted from [nasal] to [velar and/or nasal]. Interestingly, the vowel of the base form (which is usually [I]) is the least important feature for the assignment of a verb to this class. This was verified experimentally, but it is also evident in the fact that some of the newer members of this class have vowels other than [I] in their base forms, for instance, *strike*, *drag* and *sneak*.

In a feature-changing rule-based account the features of this vowel, which would be the segment undergoing the change, would be the most important part of the structural change. Bybee and Moder claim that its relative lack of importance is due to the fact that the actual class of items that forms the lexical pattern or schema consists of the past tense forms of these verbs.

This case is important for several reasons: it shows that these past tense verbs are all stored in the lexicon and associated in such a way as to form a class. It also shows that one and the same mental faculty is used for linguistic and nonlinguistic categorization, since these linguistic objects, verbs, are categorized in the same way as natural and cultural objects. It is also important that the whole word forms the basis of the categorization—including initial consonant cluster—because this demonstrates a holistic storage and processing mechanism rather than a componential or concatenative one.

Some linguists have accepted this evidence for the organization of the strong verb classes of English, while maintaining that a "symbolic rule" also exists for the formation of the regular past tense (Prasada and Pinker 1993). What is the evidence for a rule of past tense formation? The

evidence traditionally cited is the fact that children acquiring English characteristically overgeneralize the regular past tense, producing forms such as *eated*, *comed* and *breaked*. However, a recent study of past tense formation in English by Marcus et al. (1992), which takes into account over 10,000 tokens of irregular past tense produced by children after they have begun overgeneralizing, shows that overgeneralization occurs in only about 2.4% of the tokens studied. Such a small number of overgeneralizations does not support the hypothesis that regular past tense is formed differently than irregular, which is what Marcus et al. argue in spite of their data. This small percentage of overgeneralization suggests instead that regular past tense formation is just one of a number of schemas available to children. Its productivity is due to the fact that it is strongly represented, stored as it is on such a large number of verbs.

7. Some questions for the future

The purpose of this paper has been to argue for a new approach to phonology, one that is compatible with the newest cognitive and functional approaches to morphosyntax and semantics. This approach does not assume that phonological systems are symmetrical, regular, discrete, or fixed. It considers much of what phonologists analyze as rules to be the residue of historical changes. Rather than reconstructing older stages of the phonology of a language, it is proposed to study the way speakers and hearers deal with phonological regularity and irregularity and how that shapes the phonology. If lexical representations are viewed as actual cognitive entities based on the realities of perception and production, it appears that individual words must contain substantial phonetic detail. Furthermore, studies of lexical diffusion show that individual words, and thus particular sounds as parts of words, are affected by the way they are used and the frequency of their use. Thus even productive processes apply differentially to individual words, suggesting that much closer attention to phonetic detail will reveal more about lexical representations. Higher-level generalizations affecting lexical classes of words can also be shown not to be strictly rule-like in their nature, indicating that general principles of cognitive organization can be applied to this level of phonological processing.

Received 15 December 1992
Revision received 30 August 1993

University of New Mexico

Notes

1. I am grateful to the following people for comments on and reactions to this paper: Wally Chafe, Carol Genetti and Sandra Thompson.

2. Even variation theory has not fully exploited a functionalist perspective or reconsidered assumptions about underlying representations and rules, even though the data found in variation studies begs for a re-examination of all old assumptions.
3. It is well-known that the salient perceptual features by which hearers identify words are sometimes those that phonologists consider redundant or predictable, such as vowel length in English before voiced vs. voiceless obstruents, or aspiration of prevocalic voiceless stops. This mismatch between the phonologist's analysis and the perceptual facts would be remedied by including phonetic detail in lexical representations.
4. The phoneme /s/ has a frequency percentage of 7.5%, making it the most frequent consonant in Spanish. Perhaps this is one factor in its tendency to weaken and delete.

References

- Amastae, Jon
1989 The interaction of *s*-aspiration/deletion and spirantization in Honduran Spanish. *Language Variation and Change* 1, 169–183.
- Aske, Jon
1990 Disembodied rules vs. patterns in the lexicon: testing the psychological reality of Spanish stress rules. *Berkeley Linguistics Society* 16, 30–45.
- Browman, Catherine P. and Louis M. Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3, 219–252.
1990 Tiers in articulatory phonology, with some implications for casual speech. In Kingston John, and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 341–376.
- Bybee, Joan L.
1985 *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.
1988 Morphology as lexical organization. In Hammond, M. and M. Noonan (eds.), *Theoretical Morphology*. San Diego: Academic Press, 119–141.
- Bybee, Joan L. and Mary A. Brewer
1980 Explanation in morpho-phonemics: Changes in Provençal and Spanish preterite forms. *Lingua* 52, 201–242.
- Bybee, Joan L. and Elly Pardo
1981 Morphological and lexical conditioning of rules: Experimental evidence from Spanish. *Linguistics* 19, 937–968.
- Bybee, Joan L. and Carol Lynn Moder
1983 Morphological classes as natural categories. *Language* 59, 251–289.
- Dressler, Wolfgang U.
1985 *Morphology: The Dynamics of Derivation*. Ann Arbor, MI: Karoma.
1990 The cognitive perspective of “naturalist” linguistic models. *Cognitive Linguistics* 1, 75–98.
- Du Bois, John W.
1987 The discourse basis of ergativity. *Language* 63, 908–955.
- Foley, James
1977 *Foundations of Theoretical Phonology*. Cambridge: Cambridge University Press.
- Greenberg, Joseph
1966 *Language Universals*. The Hague: Mouton.

- Hooper, Joan B.
 1976a *Introduction to Natural Generative Phonology*. New York: Academic Press.
 1976b Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (ed.), *Current Progress in Historical Linguistics*. Amsterdam: North Holland, 96–105.
 1978 Constraints on schwa deletion in American English. In Fisiak, Jacek (ed.), *Recent Developments in Historical Phonology*. The Hague: Mouton, 183–207.
- Hooper, Joan Bybee
 1981 The empirical determination of phonological representations. In Myers, Terry et al. (eds.), *The Cognitive Representation of Speech*. Amsterdam: North Holland, 347–357.
- Hopper, Paul J. and Sandra Thompson
 1984 The discourse basis for lexical categories in universal grammar. *Language* 60, 703–783.
- Langacker, Ronald
 1987 *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leopold, W.
 1948 *Speech Development of a Bilingual Child: A Linguist's Record*. 4 volumes. Evanston, IL: Northwestern University Press.
- Manczak, Witold
 1980 Laws of analogy. In Fisiak, Jacek (ed.), *Historical Morphology*. The Hague: Mouton.
- Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, and Fei Xu
 1992 *Overregularization in Language Acquisition*. Monographs of the Society for Research in Child Development, Serial No. 228. Volume 57, no. 4.
- Menéndez-Pidal, Ramón
 1950 *Orígenes del español. Estado lingüístico de la península ibérica hasta el siglo XI*. Madrid: Espasa-Calpe.
- 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.
- Navarro Tomás, T.
 1946 *Estudio de fonología española*. Syracuse: Syracuse University Press.
- Pagliuca, William and Richard Mowrey
 1987a Articulatory evolution. In Giacalone Ramat, A. et al. (eds.), *Papers from the VIIIth International Conference on Historical Linguistics*. Amsterdam: Benjamins.
 1987b On natural classes. Paper presented at the Conference on Linguistic Categorization, University of Wisconsin, Milwaukee.
- Prasada, Sandeep and Steven Pinker
 1993 Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8, 1–51.
- Port, Robert F. and Michael L. O'Dell
 1985 Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13, 455–471.
- Terrell, Tracy
 1975 Functional constraints on the deletion of word-final /s/ in Cuban Spanish. *Berkeley Linguistics Society* 1.
 1977 Constraints on the aspiration and deletion of final /s/ in Cuban and Puerto Rican Spanish. *The Bilingual Review* 4, 35–51.
- Tucker, A. N. and M. A. Bryan
 1957 *Linguistic Survey of the Northern Bantu Borderland. Volume 4: Languages of the Eastern section, Great Lakes to Indian Ocean*. London: Oxford University Press (for the International African Institute).
- Vennemann, Theo
 1972 Rule inversion. *Lingua* 29, 209–242.
- Zwicky, Arnold
 1972 Note on a phonological hierarchy in English. In Stockwell, R. and R. Macaulay (eds.), *Linguistic Change and Generative Theory*. Bloomington, IN: Indiana University Press, 275–301.