18

Lexicalization of sound change and alternating environments

JOAN BYBEE

18.1 Usage-based theory

Over the last twenty years a significant functionalist trend has developed in the study of morphosyntax with the aim of explaining the nature of grammar by studying how language is used in context. The basic premise of this work is that frequently-used patterns become conventionalized or fossilized as grammatical patterns; that is, grammar is emergent from language use (Givon 1979, DuBois 1985, Hopper & Thompson 1980, 1984, Hopper 1987 and many more). Haiman (1994) has discussed the process by which repeated patterns become part of 'grammar' in terms of ritualization, showing that the effects that repeated stimuli or repeated action has on an organism—automatization, habituation and emancipation—are also operative in the process of grammaticalization or the creation of new grammar (see also Boyland 1997 for a discussion of the psychological mechanisms involved).

Some comparable research in the phonological domain has begun to appear in recent years. For instance, several studies have shown that speakers' judgements of the grammaticality of phonotactic patterns is based on the frequency of consonant and vowel combinations actually occurring in the language (Pierrehumbert 1994, Frisch 1996). In addition, speakers' ability to access the lexicon may involve a complex interplay between the frequency of words and the number and frequency of words with similar phonological shape (Pisoni, Howard, Nusbaum, Luce & Slowiaczek 1985). Connectionism offers the possibility of formally modeling the effect of use on mental representations of language and such models have been tested in the phonological and morphological domains (for example in Daugherty & Seidenberg 1994, and Dell 1989.)

The current paper is intended to contribute to the general effort to show how language use can be recruited to help explain some well-known properties of phonological patterns. The data examined, which is from phonological variation,

also provide some evidence for the size and nature of the phonological memory representations for words and phrases.

First, I present evidence that many, if not all, sound changes progress in lexical items as they are used, with more frequently-used words undergoing change at a faster rate than less-frequently used words. Then I examine 'alternating environments'—cases in which a sound in a particular word or morpheme is sometimes in the environment for the change to take place and sometimes not. In cases in which the targeted sound is at the edge of a word, the change can go through even where the sound is not in the appropriate phonetic environment and thus no alternation is produced. In such cases, we have evidence for the restructuring of the lexical representation of the word. On the other hand, when the alternating environment is inside of a word, the change can be retarded even in the appropriate environment, but eventually an alternation can be created, showing, again, restructuring of the lexical representation of the word. I argue that lexical representations are restructured gradually on the basis of actually occurring variants of a word, and that postulating words and frequent phrases as the units of representation explains the development of word-level phonology. In addition, it will be argued that reference to the frequency with which words begin in consonants explains why final word boundary often conditions changes as though it were a consonant.

18.2 The frequency effect on sound change

One of the aims of this paper is to explore from a phonological perspective the size and nature of storage and processing units. I will present evidence that words and often longer units, such as frequent phrases, are the units of lexical storage. For the moment, however, I assume that words are the units of lexical storage. It is reasonable to assume that lexically-stored words are in many ways like other mental records of a person's experience. First, there is no reason to believe that these memorial records have details and predictable features abstracted away from them (Langacker 1987, Ohala & Ohala 1995), and second, it is reasonable to believe that new experiences are categorized, to the extent possible, in terms of the already stored record of past experiences (see Klatzky 1980).

Each use of a word requires retrieval by the speaker and a matching of the incoming percept to stored images by the hearer (and the speaker, who is monitoring his or her own speech). My thesis in this paper is that the act of using a word, either in production or perception, has an effect on the stored representation of the word. We already know this is true in terms of the degree of entrenchment of a word (or the resting level of activation): high frequency words have stronger representations which make them easier to access, more resistant

to change on the basis of other patterns, and more likely to serve as the basis for the creation of new forms (Bybee 1985).

In addition, certain levels of use affect the stored representation of words by actually changing their shapes. That is, along with the entrenchment effect of frequency, there is also an automation effect: words and phrases that are used a lot are reduced and compressed. This effect is very salient in grammaticizing phrases (such as *going to* becoming *gonna* and *want to* becoming *wanna*) and more conventionalized contractions (such as *won't* and *didn't*), but it also occurs in a more subtle form across the lexicon when a sound change is taking place. Sound changes (phonetically motivated changes, which are usually the reduction of the magnitude of gestures or retiming of gestures Browman & Goldstein 1992a) tend to be phonetically gradual and also lexically gradual: high frequency words undergo change at a faster rate than low frequency words. The effects of frequency in the diffusion of a sound change through the lexicon have been shown for vowel reduction and deletion in English (Fidelholz 1975, Hooper 1976b), for the raising of /a/ to /o/ before nasals in Old English (Phillips 1984), for various changes in Ethiopian languages (Leslau 1969), for the weakening of stops in American English and vowel change in the Cologne dialect of German (T. Johnson 1983), for ongoing vowel changes in San Francisco English (Moonwomon 1992), and for tensing of short *a* in Philadelphia (Labov 1994:506–7). In a recent paper, I have shown that there is also a frequency effect in the application of t/d-deletion in American English (Bybee 1998b). Deletion occurs more in high frequency words, including of course monomorphemic nouns and adjectives, but also regular past tense verbs, a point to which I will return later.

My interpretation of the frequency effect in the diffusion of sound change (following Moonwomon 1992) is that sound change takes place in small increments in real time as words are used. The more a word is used the more it is exposed to the reductive effect of articulatory automation. The effects that production pressures have on the word are registered in the stored representation, probably as an ever-adjusting range of variation. Thus words of higher frequency undergo more adjustments and register the effects of sound change more rapidly than low frequency words.

The frequency with which a word is subject to the ravages of articulation is not the only factor that encourages sound change. We also have to take into account the fact that certain speech styles allow more reduction and compression than others. In particular, casual speech among familiars typically shows more reduction. Thus words that are used in casual situations will also undergo change at a faster rate (D'Introno & Sosa 1986). Of course, these words are also likely to be those that are of higher frequency overall.

Another factor affecting reduction is the status of the word within the discourse. Fowler & Housum (1987) found that the first use of a word in a spoken

text was longer than in subsequent uses. This means that speakers articulate more clearly in the first use of a word, where identification by the hearer might be more difficult, and then allow the reductive processes to apply later when identification by the listener is aided by the context and the fact that the word has already been activated. In fact, speakers may use reduction to indicate that a referent is *not* new, but rather one that has already been accessed in the discourse. Words that are used more often within a text are produced in reduced form more often. If the produced form affects the stored form, then words that are repeated more often in a discourse will reduce at a faster rate than words that are repeated less often.

18.3 Exemplar based representations

The account of phonetically gradual lexical diffusion of a sound change given in the preceding section requires a model of memory storage for linguistic units based on actual tokens of use. Each experience of a word is stored in memory with other examples of use of the same word. These memories of specific tokens are organized into clusters with more frequently-occurring exemplars, and tokens that share many properties with high frequency exemplars are treated as more central, while less common or more deviant tokens are treated as more marginal. Thus linguistic experiences are categorized in the same way as other types of perceptual experiences. Rather than conceiving of stored representations as abstractions from the phonetic tokens, representations are considered to be the result of the categorization of phonetic tokens. This proposal, which will be referred to as 'the exemplar model', adapts proposals made by Miller (1994) for phonetic segments and K. Johnson (1997) for larger units. Similar arguments for phonological representations have been made by Hooper (1981) and Cole & Hualde (1998). Note that this model does not distinguish between phonetic and phonemic features in lexical representation (see Steriade, this volume). Further implications of this model for sound change will be discussed in the next sections.

18.4 Alternating environments

Given that produced tokens affect stored representations, what would happen when a word or morpheme occurs in different environments, such that it is subject to a change in one environment but not in another? In the case of such 'alternating environments' (as Timberlake (1978) calls them) two or more different surface forms map onto a stored form. How are such alternate mappings resolved?

Here I approach this question by examining cases of sound change in progress. It is necessary to distinguish the phonetic variation that goes on while a

sound change is in progress from the conventionalized alternations that can eventually arise from such sound change. By 'alternation' I mean that a word or morpheme has two or more variants that are not phonetically continuous nor variable, but rather constitute discrete alternants conditioned by specific phonological, grammatical or lexical contexts. An alternation, then, roughly corresponds to the level of variation generated by a classical phonological rule. By studying the conditions under which such alternations are conventionalized and the conditions under which they are not, we learn something about how variants of words and morphemes are organized in memory.

The study of alternating environments in cases of sound change in progress reveals that the outcome differs according to whether it is a word or a morpheme that is in the alternating environment. When the same morpheme is in an alternating environment in different words, a change is retarded even in the conditioning environment (the Timberlake Effect, see below), but an alternation can eventually arise. When the alternates are in two forms of the same word, alternations arise only under special conditions, but ordinarily only one alternate survives. I will argue that the differential behavior of morphemes and words with respect to sound change in progress provides strong evidence for the stored representation of words and frequent phrases.

18.4.1 Sound changes at word boundaries

The fact that phonological phenomena occurring around word boundaries often have a different effect from phenomena inside of words can be attributed to the use of forms in context and the way this usage affects stored representations. As an example consider the well-studied variable phenomenon of syllable-final *s*-aspiration and deletion in Latin American Spanish. A syllable-final /s/ before a consonant is subject to loss of its lingual articulation, resulting in what has been termed 'aspiration' or a period of voiceless frication, which itself is eventually subject to loss. This change can affect a word-internal /s/ as well as a word-final /s/ before a consonant (see examples in (1)). As this change is ongoing in many dialects of Latin America, we can compare two dialects that appear to be at different stages in order to observe differences in the effect that the sound change has on word-final vs. word-internal /s/.

The examples in (1) illustrate the variation found in /s/ in the different phonological environments where /s/ deletes or aspirates in Latin American Spanish (Terrell 1978). Non-phonological factors involved in this variation include the sex and age of the speaker, lexical factors (see below), speaking rate and register.

Consider now Table 18.1, which shows the rate of deletion in different contexts in Argentine and Cuban Spanish. These tables generalize over thousands of tokens produced by dozens of subjects and obscure some important factors,

including some lexical factors that will be mentioned below. For present purposes, however, let us focus on one difference between the two dialects.

(1) a. ___C: word-internally before a C

feli*h*mente	'happily'
e*h*tilo	'style'
dentis*t*a	'dentist'

b. ___##C: word-finally before a C

o se traen animale*h* finos	'or fine animals are brought'
haya muchos temas	'there are many themes'
su∅ detalleh	'his details'

c. ___##V: word-finally before a V

y mientra*h* esa sonoridad así	'and during this voicing thus'
no vas a encontrar	'you are not going to find'

d. ___//: before a pause

en momentos // libre*h* //	'in moments, free'

Table 18.1 *Gradual lexical restructuring: Spanish s-aspiration (Terrell 1977, 1978, 1979, Hooper, 1981)*

Argentine Spanish

	s	h	Ø	tokens
___C	12%	80%	8%	4150
___##C	11%	69%	20%	5475
___##V	88%	7%	5%	2649
___//	78%	11%	11%	2407

Cuban Spanish

	s	h	Ø	tokens
___C	3%	97%	0%	1714
___##C	2%	75%	23%	3265
___##V	18%	48%	34%	1300
___//	61%	13%	26%	1776

The Argentine dialect can be considered to be in an early stage of the implementation of this change relative to the Cuban dialect. The important point to note about Argentine Spanish is that at this stage, the word-boundary appears to have little effect. Looking at the first column, we see that the maintenance of /s/

is largely predictable from the phonetic environment: before a C the maintenance of /s/ is at 11–12%, with no significant difference between word-internal and word-final /s/. The maintenance rate at the end of a word before a V (__##V) is at 88%, since a following vowel does not constitute the appropriate environment for the change. Now compare Cuban Spanish: the fact that this dialect is in a more advanced stage of the change is indicated by the fact that the rate of maintenance of /s/ is only 2–3% before a consonant. Again, the presence of the word boundary makes little difference. The major difference between Argentine and Cuban Spanish for our purposes is in the use of /s/ at the end of a word before a vowel. In Cuban Spanish aspiration and deletion are common in this context, with maintenance of /s/ down from 88% in Argentine to 18% in Cuban, even though it is not the appropriate phonetic context for the change and no such change is going on with word-internal /s/ before a vowel.

What would cause a phonetic change to occur outside of its phonetic environment? Obviously it is the position of the /s/ at the end of the word that is crucial: in this alternating environment, the /s/ is sometimes in the position for aspiration and sometimes not. In fact, more than half the time, word-final /s/ is in the environment for aspiration as shown in Table 18.2.

Table 18.2 *Percentage of occurrence of word-final /s/ before a consonant, vowel and pause*

Argentine		Cuban	
C	52.0%	C	51.5%
V	25.1%	V	20.5%
//	22.9%	//	28.0%

If, as I hypothesized above, sound change affects stored representations incrementally each time a word is used, the use of a word-final /s/ before a consonant will have some effect on the stored representation, especially if there is only one representation per word. Thus the /s/ will gradually decay in the stored representation and the reduced form of that consonant will eventually appear even before vowels, as in Cuban Spanish.

Note that this case exhibits the classic feature of a final word boundary behaving like a consonant. The figures in Table 18.2 provide us with a usage-based explanation of this common phenomenon: if usage affects stored representations and if twice as many tokens begin with a consonant than with a vowel, phonetic changes conditioned by a following consonant will also take

place at a word-boundary, though at a slower rate than word-internally, where the environment does not alternate.

Ordinarily, alternations do not develop where the conditioning is across a word boundary. This fact gives rise to the notion of 'word-level phonology'—that is, the fact that most alternations occur within words. The explanation being investigated here is that ordinarily there is only *one* stored representation for each word. Where variation arises during change in progress, the variation is resolved in terms of one variant or the other. The exceptions to this arise only in the case of frequently used phrases, to which we will return shortly.

First let us consider how variation at the word level is represented and how cases of sound change in an alternating environment would eventually be resolved. In the exemplar model described above, the representation of a word is a cluster of actually-occurring tokens, with more frequent tokens accumulating greater weight or strength. Thus each word has its own range of variation dependent upon its frequency and the contexts in which it is used. When little or no sound change is affecting a word, the range of variation in the tokens may be small and relatively stable. During change, however, the range of variation increases and the center of the cluster gradually shifts.

When the same word occurs in both an environment that conditions a change and in a non-conditioning environment, as in the Spanish *s*-aspiration case, the cluster for a word may divide into two (or more) subclusters, each one with a strong center of high frequency tokens. In this case, each subcluster is associated with one environment—the word-final [s] tokens with the environment before a vowel and the word-final [h] tokens with the environment before a C. It appears that such a situation is unstable when the environment is not also part of the representation, because it tends to be resolved in favor of one variant for all environments. That is, the most frequent variant, the weakened consonant, [h], wins out and tends to be chosen even in contexts before a vowel.

In contrast, when the environment is part of the stored unit, an alternation can be established, in the sense that the [s] can remain before a vowel. This happens in frequent phrases. For example, Terrell (1986:129) reported that one illiterate speaker of Dominican Spanish used word-final /s/ only twelve times out of 443 words with orthographic /s/ in a taped interview. All cases involved grammatical morphemes in set phrases: /s/ is used four times with a plural definite article followed by a stressed vowel (e.g. *las otras* 'the others (fem.)', *las únicas* 'the only ones (fem.)' etc.), three times in the phrase *más o menos* 'more or less', in the names of two ball teams, *Las Estrellas* 'The Stars' and *Las Aguilas* 'The Eagles', in the phrase *es igual* 'it's the same' and in the words *ellos* 'they (masc.)', *tres* 'three'. In the last two cases Terrell does not report what word followed, but it is highly probable that it is a high frequency vowel-initial verb after *ellos* and a high frequency vowel-initial noun after *tres*. The conclusion is,

then, that in frequent phrases that constitute lexical units, the /s/ is preserved before a vowel, even though in other contexts the word may have lost its /s/.

The Dominican case is extreme, as the loss of word-final [s] is quite advanced. In the Cuban case, which is not so advanced, I extracted the examples of orthographic *s* before a vowel (as transcribed by Terrell) from the interview of one speaker. In this interview there were 77 opportunities for the speaker to produce an [s] before a vowel; only twelve of these were transcribed as [s]. (This constitute 15%—a number close to the 18% reported for all speakers.) These cases of word-final [s] before a vowel were:

(2) a. the phrase *las once* meaning 'eleven o'clock'

b. four cases of a quantifier with the noun *años* 'years'
seis años 'six years', *tres años* 'three years', *muchos año(s)* 'many years'

c. the adjective-noun phrase *determinados época* 'certain times'

d. five occurrences with the third singular copula *es*, followed by a definite or indefinite article: *es al medio día* 'it's at noon', *es el centro* 'it's the center', *es el sureste* 'it's the southeast', *es en lo general* 'that's in general', *es un...* (pause) 'it's a...'

The phrase in (a) is a standard phrase used in speaking about the time of day. The phrases in (b) and (c) are all quantified nouns designating temporal units. The [s] on the quantifier is retained before the frequently-used vowel-initial noun *años*. The cases in (d) are all examples of the third singular copula *es* followed by a grammatical morpheme—either a preposition or an article. All of these sequences (with the possible exception of (c)) occur commonly enough to constitute processing and storage units. The [s] is retained in these phrases because the phrases are acting like words, and the [s] is in a sense 'word-internal' in these cases.

Thus my hypothesis is that words and frequent phrases are storage units and that ordinarily there is only one representation per word, so that variations in the form of a word are normally reconciled to a single form, and no alternation is created through sound change. Exceptions to this occur when a word is used in high frequency phrases and/or phrases involving grammatical morphemes, such as pronouns and articles. This hypothesis makes strong predictions about the conditions under which sandhi phenomena will develop. It predicts that sandhi processes will only occur in phrases of high frequency and most commonly in those involving grammatical morphemes, or other high frequency words. This prediction is borne out by the most famous cases of sandhi, such as French liaison (Tranel 1981).[1] The hypothesis also predicts that cases of reduction restricted to certain 'syntactic' environments, such as English auxiliary contraction and the reduction of *don't* will occur only in the most frequent contexts in which

the form appears. This prediction is also borne out where it has been tested (Bybee and Scheibman, to appear; Krug, to appear. See also Bybee, 1998a.).

18.4.2 Sound change inside of words

Of course, alternations do develop inside of words. A morpheme inside a word may undergo phonological change, producing a new allomorph, and thus an alternation. This fact follows from the hypotheses presented above: if sound change permanently affects stored units, and words—even morphologically complex ones—are the units of storage, then the same morpheme in different words will take on different phonological shapes.

Further evidence for the hypotheses developed here is the fact that the effect of an alternating environment inside a word is very different from the effect across word boundaries. Inside a word, a variable process never applies outside of its phonetic environment (as, say, the aspiration of /s/ in Spanish occurs even ___##V). Instead the effect is the reverse: there is evidence that a change can be retarded even in its phonetic environment if it occurs in a morpheme which also has alternates which appear outside of the phonetic environment. Timberlake (1978) has made this point by presenting examples of changes that progress faster in uniform environments and are retarded in alternating ones. I thus dub this the 'Timberlake Effect'. Consider the following examples from the Mazowian dialect of Polish (Timberlake 1978: 313–314). Transcriptions are from Timberlake and his sources.

In root-internal position velars are always more or less palatalized before /i/ (the symbol [ʲ] indicates full palatalization and ['] indicates partial palatalization):

(3)	kʲil	'stick'
	skʲiba	'ridge'
	gʲipsu	'gypsum'
	k'ɪlômetr	'kilometer'
	g'ɪrie	'bends'

In alternating environments, represented here by position before a suffix, stops are palatalized before /i/ in only about half of the recorded examples, in (4).

Where the velar is stem-final, it will sometimes occur in a palatalizing context and sometimes not. This is the reason, according to Timberlake, that the palatalization process is retarded even when the conditioning is present.

The same effect may be observed with affixes. Affixes, such as English Past Tense [t] or [d], are in an alternating environment, because they may occur after vowels, as in *played,* or *tried*, and they may occur after consonants, as in *walked* or *learned.*

(4)				
	prog'ɪ	'hearths' (Gen. Sg.)		
	drog'ɪ	'roads' (Nom. Sg.)		
	durak'ɪ	'beets'		
	jarmak'ɪ	'fairs'	morg'ɪ	'acres'
	rog'ɪ	'horns'	gruskɪ	'pears'
	kackɪ	'ducks'	zmarsckɪ	'wrinkles'
	drugɪ	'other'	robakɪ	'worms'

The suffix is only in a context for deletion when it is preceded and followed by a consonant. A monomorphemic word also has an alternating following environment, but its preceding environment is uniform. It could be for this reason that Past Tense [t] and [d] delete less often than [t] or [d] in monomorphemic words, as shown in Table 18.3 (data from Bybee 1998b).[2]

Table 18.3 *Rate of deletion for Regular Past tense compared to all other words of comparable frequency (403 or less)*

	Percentage deletion
All words	45.8%
-ed verbs	22.6%

In contrast, sound changes in affixes in internal and non-alternating environments go through before all others: deletion of Spanish /d/, which is allophonically [ð], is much more advanced in the Present Participle *-ado* than in other contexts (excepting high frequency words, such as *lado* 'side') (D'Introno & Sosa 1986). Similarly, a very early instance of the loss of intervocalic /d/ occurred in the Second Plural morpheme *-ades*, which became *-ais* in Old Spanish.

These examples show that the implementation of a sound change in particular words (that is, its lexical diffusion) depends heavily on the contexts in which the sound is used. Since I have been arguing that the unit that serves as the context for a sound as it undergoes change is the word, then we must now consider how to account for the fact that the environment of a morpheme in one word affects the rate of change for the same morpheme in a different word. To understand this issue, we must understand the nature of morphological relationships, a matter to which we now turn.

18.5 A network model

In various works (Bybee 1985, 1988, 1995) I have proposed that the lexicon is organized into a complex set of relations among words and phrases by

connections drawn among phonologically and semantically similar items. Parallel phonological and semantic connections constitute morphological relations if they are repeated across multiple pairs of items. As Dell (this volume) points out, morphological relatedness is the joint effect of the organization of words into phonological and semantic neighborhoods.

In this model, the relations between base and Past Tense forms of English verbs are diagrammed as in Figure 18.1, where semantic relations are not explicitly shown, and where relations of similarity (rather than identity) between segments is shown with broken lines.

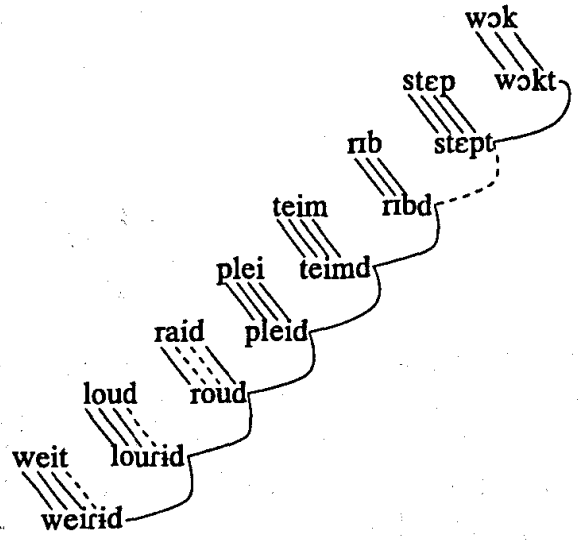

Figure 18.1 Model of relations between English Past Tense forms and their bases.[3]

Affixes are not explicitly listed in storage but emerge from sets of connections made among stored words and phrases. The very high type frequency of the regular English Past Tense strengthens its representation in memory and makes it highly productive. It can then apply to verbs whose Past Tense forms are not accessible because they have never been encountered or are of such low frequency as to not be easily accessible.

Past Tense constitutes a category, but not one that can be accessed independently of a particular verb, because it is a category to which verb forms may belong or not belong. How then do individual tokens of the Past Tense suffix relate to one another? This is, of course, an empirical question and here is the evidence we have so far.

First, instances of the suffix attached to a verb are affected by the token frequency of the whole verb form: the rate of deletion for a final [t] or [d] on a Past Tense verb is affected by the frequency of the form, as shown Bybee (1998b). In that study, Past Tense forms with a frequency in Francis & Kučera (1982) of 36 or greater are considered high frequency and those with a frequency of less than

36 as low frequency, following a suggestion by Stemberger & MacWhinney (1988), who establish that the mean frequency of inflected verbs in Francis and Kučera is 35. Using this cut-off point, Bybee (1998b) finds that there is a significant difference between high and low frequency verbs in the position for deletion. See Table 18.4.

Table 18.4 *The effects of word frequency on t/d-deletion in regular Past Tense verbs (non-prevocalic only).*

	Deletion	Non-deletion	% Deletion
high frequency	44	67	39.6%
low frequency	11	47	18.9%

χ^2: 5.00313, $p < .05$, df = 1

Second, in the data we cited above, the overall trend for Past Tense [t] and [d] is that they delete *less* often than [t] or [d] in monomorphemic words.

A related third point is that Lociewicz (1992) has shown that not only are monomorphemic [t] and [d] shorter than Past Tense [t] and [d], but high frequency Past Tense [t] and [d] are shorter than low frequency [t] and [d]. Lociewicz proposes to account for her data by a dual-access model in which high frequency morphologically complex forms are stored and retrieved as wholes, while low frequency forms are composed by adding the suffix to a base form using a schema. But this model would predict the same rate of deletion in high frequency regular Past Tense forms as in monomorphemic forms of comparable frequency, and this prediction is not borne out by the data. Rather, in the data used in Bybee (1998b), the rate of deletion for all words with frequencies of 36–403 was 54.4%, while the rate for Past Tense forms of the same frequency was 39.6%. Thus we must posit that the Past Tense in low frequency verbs, which is longer and phonetically fuller, can have some effect on the Past Tense suffix on high frequency verbs, which is shorter and more prone to deletion, but not as short and prone to deletion as the [t] and [d] of monomorphemic forms. Thus the fuller form of the suffix on low frequency verbs has some impact on the suffix on other verbs.

A proposal that would account for all the facts accumulated so far would be one in which the Past Tense suffix forms a category, perhaps in terms of a set of related schemas, not separate from the forms it describes, but emergent from them.

[___ [+voice]]$_{verb}$	d]$_{past}$
[___ [–voice]]$_{verb}$	t]$_{past}$
[___ [t/d]]$_{verb}$	id]$_{past}$

Different tokens of Past Tense contribute to the formation of the category and the data suggest that the category itself influences the shape of the productions. Since the occurrences of cases in the environment for deletion are many fewer than the cases that are protected from deletion (by surrounding vowels), the deletion is retarded in all members of the category Past Tense verb.

Now we return to the Polish examples, in which an alternating environment for a *stem* seems to slow down a change in progress elsewhere. The inflected words affected by the change are embedded in a dense network of related forms, which includes the connections shown in Figure 18.2.

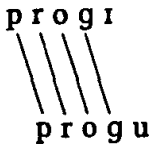

Figure 18.2 Lexical connections between two instances of the same stem.

The closely related forms *progu* and *progi* are separate words and each has a representation, however the shared stem also forms a category. As the palatalization of [g] before a high front vowel proceeds, the range of variation for the stem *prog-* begins to widen. The Timberlake Effect is evident now, as the center of the category is affected by the variants in the non-palatalizing environment. Since the two variants occur in different words—their environments are part of the representation—they can eventually diverge into two forms of the stem, creating an alternation.

This account of the Timberlake Effect makes predictions about the circumstances under which the effect will be the strongest. A change will be retarded most noticeably in an alternating environment when the alternates that are not in the environment to undergo the change are the most frequent—either there are more conditions in the paradigm in which the change does not take place, or the environments that do not condition a change occur in the unmarked, or most frequent categories. Furthermore, it is less likely that the Timberlake Effect will be observed in high frequency paradigms in which individual forms have a greater lexical strength (accessibility) and weaker connections with related words (see Bybee 1985).

Now compare again these word-internal cases to those where change is occurring at a word boundary. The word may for a time have multiple variants, suggesting either a range of variation in the changing segment, or even multiple

representations for a single word. However, in this case the tendency is to resolve the variation in favor of a single form for each word, except in the case of high frequency or grammatical words. It appears that the cases in which distinct alternates become established are just those cases in which the conditioning environment is registered in storage with the alternating item. Thus in the case of *progu, progi* each variant can be registered because we are dealing with two different (although related) words, one of which consistently has the palatalizing environment and one which does not. Similarly in frequent phrases, such as *muchos año(s)* the [s] preceding the vowel may be preserved (as though it were word-internal) because the conditioning vowel occurs with it in storage and processing. Other instances of the same word may occur without the [s], as [mučoh] or [mučo]. Such variation would not necessarily exist indefinitely. Unless one variant is in a highly entrenched phrase, the variation is likely to be eventually leveled out.

Thus by registering words in the lexicon and establishing connections among them, we are able to account for the two different effects on sound change of alternating environments inside of words and across word boundaries.

18.6 Lexical phonology

Some of the effects of variable processes that I have discussed above have been addressed by Guy (1991a, 1991b) in the context of Lexical Phonology. This proposal is relevant here, even though I will argue that it does not work for all the cases at hand, because it incorporates the notion argued for here, that some words behave as if variable processes have applied to them more than once. Guy proposes that variable rules may apply cyclically and at all levels of a Lexical Phonology, and offers an account for the variation in t/d-deletion which is conditioned by the morphological structure of the word. The facts are as follows: on average, the highest rate of deletion of /t/ or /d/ takes place in monomorphemic words, the next highest rate in pasts with vowel changes (such as *slept, left, told*), and the lowest rate in regular past tense forms.

(5)	[perfect]	[sleep]	[check]	
				Level 1
	may apply	———	———	t/d-deletion
		[slept]	———	irregular inflection
	may apply	may apply		t/d-deletion
				Level 2
			[checkd]	regular inflection
	may apply	may apply	may apply	t/d-deletion
	3 chances	2 chances	1 chance	

Guy proposes derivations, as shown in (5), in which t/d-deletion applies variably, and at the same rate all the time, but it has three chances to apply to monomorphemic words, two chances to apply to vowel-change pasts and one chance to apply to regular pasts.

The Lexical Phonology approach also works for Timberlake's first example, as shown in (6), where palatalization may be thought of as increasing incrementally each time the palatalization rule applies. This derives the effect of having greater degrees of palatalization in uniform than in alternating environments.

(6)	[kij]	[prog] i	
			Level 1
	[k'ij]	———	Palatalization
			Level 2
		[progi]	Regular Inflection
	[kʲij]	[prog'ɪ]	Palatalization

The advantage of the Lexical Phonology approach is that it does recognize that fairly low-level variable phonology is deeply entwined with the lexicon and morphology. It also suggests that the greater progress of a sound change can be attributed to more applications of the 'rule.' The problem with it is that it makes incorrect predictions in some cases, and it cannot deal at all with frequency effects across lexical items.

Consider first another case of an alternating environment which Timberlake describes. Timberlake brings up this case to show that it is the alternating environment itself, and not the morpheme boundary, that causes the retardation of change. In this case, also from Polish, there is palatalization of /k/ quite regularly and over a large geographic area before /e/ in a uniform environment, such as in the word 'sausage', [kʲeu̯basa], but palatalization in [u̯okeć] 'elbow' (Nom. Sg.) is less developed and more geographically restricted, due to its alternation with the form [u̯okća] (Gen. Sg.).

The Lexical Phonology analysis of this case predicts palatalization outside the environment for [u̯okća] and advanced palatalization in [u̯okeć], as shown in (7):

(7)	[u̯okeć]	[u̯okeć] a	
			Level 1
	[u̯okʲ'eć]	[u̯ok'eć] a	Palatalization
			Level 2
	[u̯ok'ća]		Vowel Deletion
	* [u̯okʲeć]	* [u̯ok'ća]	Palatalization

The level-ordering approach could be made to work in this case by formulating the *e/zero* alternation as a vowel insertion rather than a deletion. However, most treatments of Polish regard it as a deletion, and such a change would not solve all the problems with this approach.

The Lexical Phonology approach also makes incorrect predictions concerning sound changes affecting affixes, where the context for the change is non-alternating. Earlier I mentioned the case of Spanish /d/, which is disappearing faster in the non-alternating environment of the Past Participle suffix, *-ado*, than in stems. Stems would be subject to the reduction or deletion rule both at Level 1 and at Level 2, but *-ado*, since it is a part of regular inflection, would not be available to undergo the rule until Level 2. It would thus have *less* reduction and deletion than a stem, rather than more. This example shows that the Lexical Phonology approach is fundamentally the wrong approach, for it is a fact of usage, not structure, that is accelerating the change: since the /d/ in /ado/ is in the context for reduction and deletion no matter what verb it is added to, and since many verbs with this suffix are of very high frequency, there is a frequency effect to accelerate the change and nothing to impede it.

Finally, Lexical Phonology, as a theory of structure and not a theory of usage, cannot account for the frequency effects demonstrable in the lexical diffusion of sound change. In Bybee (1998b) I have shown that t/d-deletion occurs more often in words of higher frequency. This is true of all of the 2000 tokens studied; this relation also holds when nouns and adjectives, semi-weak Past Tense verbs, and regular Past Tense verbs are considered as well. Since the semi-weak Past Tense verbs are all of high frequency, frequency of use alone can account for their higher rate of deletion over the regular Past verbs. I conclude, then, that variable rates of phonological change are the product of usage, not of structure.

18.7 Conclusions

The evidence discussed in this paper bears on two issues regarding the nature of stored memory for linguistic forms. First, the minimal unit of independent storage is the word, which is also the minimal unit of production since smaller units cannot be used in isolation. I hasten to add, however, that that does not mean that other much longer sequences are not stored and processed as wholes. Here we have seen evidence that frequently-used phrases behave like single processing units (just as words typically do) in that they preserve segments that might otherwise be lost at word edges. In other papers I have argued for a highly redundant storage mechanism that includes specific instances of phrases and clauses as well as more generalized constructions as storage and processing units (Bybee 1998a, Bybee & Scheibman, to appear).

The view of sound change as affecting *sounds in words* according to their context of use, allows us to understand why most phonological alternations

occur at the word level: alternations can only be established in cases in which the conditioning environment is present in the storage and processing unit. Words or other units that occur in alternating environments that are not part of the stored unit will not have variants, but rather will resolve any variation in favor of one form or the other. This proposal also allows us to make interesting predictions about the development of liaison or sandhi phenomena. Conventionalized alternations across traditional word boundaries indicate that at least one alternate is part of a larger stored unit. Thus such liaison alternations can be used to study the nature and size of storage units. Finally, the view of sound change as affecting sounds in words provides an account of the different effects of alternating environments inside of words and across word boundaries.

The second major aspect of the model presented in this paper is that sound change has an immediate and permanent effect on stored representations. This view contrasts with the generative and structural view that underlying representations remain fixed and sound change is 'rule addition'—nothing more than a change in the phonological component. The evidence that sound change has an immediate effect on the lexicon is that words change gradually and at different rates according to their token frequency, even while a 'rule' is still 'variable'. The evidence that such change is permanent is the fact that old underlying forms *never* resurface, even when the 'phonological rule' becomes unproductive (see Cole & Hualde 1998 for more evidence on this point). Instead the progress of a change is inexorably unidirectional in both a phonetic and morpho-lexical sense. In the phonetic sense, we see the unidirectionality in chains of reduction and assimilation changes, such as those shown in (8), where one change builds on the other and continues its direction.

(8) $t \rightarrow d \rightarrow ð \rightarrow Ø$
$s \rightarrow h \rightarrow Ø$
$k' \rightarrow k^j \rightarrow c$

If stored items are changed gradually and the motivation for increased automation remains fairly constant, then the continuous nature and strict directionality of such changes is predicted. If sound change were 'rule addition', there would be no explanation for why e.g. after 'adding the rule' $d \rightarrow ð / V\_\_V$, a language would go on to 'add the rule' $d \rightarrow Ø / V\_\_V$.

Inexorable unidirectionality is also apparent in the morphologization and lexicalization of the results of sound change. While no one disputes that morphologization eventually takes place, I have shown here and elsewhere that involvement with the lexicon and grammar occurs very early (Hooper 1976a, 1981, Bybee 1998b). Examples given above are the word frequency effect in lexical diffusion, the lower rate of deletion of morphemic /t/ and /d/ in American English, the appearance of aspiration for earlier /s/ before a vowel at the end of a

word in Cuban Spanish and the examples described as the Timberlake Effect. Once involved with the lexicon and morphology, alternations become more and more entrenched and can only be undone by the strong pattern pressures we know as analogical leveling.

The two hypotheses of words as storage units and the immediate and permanent effect of sound change on words explain why most phonological alternations occur at word level, that is, why word boundaries block phonological 'rules' and morpheme boundaries do not. I have also shown here that the tendency of final word boundaries to act like consonants follows from these hypotheses and from the fact that the segment most frequently following a final word boundary is a consonant.

The larger theoretical message is that use impacts representation, a point often made in studies of the discourse origins of syntax and a point that is also being made by connectionist modelers of language. As I have argued here, many cases of what was earlier postulated to be structural turn out to be derivable from the way language is used. I also see many instances where a careful look at use brings to light new data that was ignored before. I suspect that a usage-based perspective will be very productive in generating new questions and new answers in phonology.

Notes

1 In fact, the conditions under which French word-final consonant appear as liaison consonants are strikingly similar to the conditions under which [s] is retained in Caribbean dialects of Spanish (Terrell & Tranel 1978).

2 These data include only cases where a consonant preceded the final /t/ or /d/.

3 Figure 18.1 highlights the emergent morphological relationships that exist between English Past Tense forms and their bases. Of course, other relationships exist, but are not shown here for reasons of simplicity.