Patterns of lexical diffusion and articulatory motivation for sound change

Joan Bybee
University of New Mexico

Patterns of lexical diffusion can serve as important diagnostics for the source of sound change. The most common lexical diffusion pattern for sound change is from high frequency words to low frequency words. This pattern is consistent with an articulatory source for change, as compared to a perceptual source. As perception must also play an important role in change, a model that includes the interaction of articulatory and perceptual change is proposed. Another set of changes that are unlikely to have an articulatory source are discussed and their properties are compared to those with articulatory motivation.

1. The source of sound change

This chapter seeks to contribute to a typology of sound change based on the causes and mechanisms of change. The theory to which it contributes would explain both why and how sound change occurs, seeking where necessary different mechanisms underlying different kinds of change. The view developed here is situated within Usage-Based Theory (Barlow & Kemmer 2000; Bybee 2001, 2002b, 2010; Croft 2000; Phillips 2006; Bybee & Beckner 2010) in that language change is attributed to processes applying in language use; it is not an isolated phenomenon, but occurs as real speakers and listeners use words, phrases and constructions of their language. Thus the typology suggested here involves reference to features of phonetics, phonology and lexicon.

There is some agreement in the literature that sound changes that are found independently in diverse languages are based on universal phonetic biases and constitute the largest portion of known sound changes (Ohala 1993). Most of the current chapter will be focused on explanations for this body of phenomena, which can be called ‘sound change’. However, other types of phonological changes such as dissimilations and phonetically abrupt changes do occur and for this reason it is important to identify the various properties of individual changes that can help to
distinguish among them and to search for causes, which are likely to be different where different properties occur.

Some of the properties of change that can aid in identifying mechanisms of change include the phonetic properties of the change such as its trajectory, whether the change is phonetically gradual or abrupt and whether or not the change has been documented across languages. Another set of properties concerns the outcome: whether or not novel segments are created and whether or not the conditioning environment is eventually lost. These properties will be mentioned in the following discussion, but the focus of our interest will be on what patterns of lexical diffusion can tell us about the source and causes of sound change. To this end, we consider the type of lexical diffusion. If it is based on word frequency, we consider the direction—from high frequency to low frequency words or the reverse. The final consideration is whether in the end the outcome of the change is lexically regular or not.

This chapter argues for lexical diffusion as a diagnostic for the mechanisms underlying changes of different types (Hooper 1976; Phillips 1984, 2006; Bybee 2000, 2001). The argument is that the most common type of lexical diffusion—from high frequency words to low frequency words—suggests an articulatory mechanism for sound change, which is the result of neuromotor automation of articulatory routines. The proposal is situated within an exemplar model which integrates gradual lexical diffusion with more general articulatory change that results in lexical regularity. The hypothesis of an articulatory basis for sound change is compared to proposals for perceptually-based sound change and it is argued that articulatory factors initiate sound change with perceptual change playing a secondary role. Other types of change, in which perception may play a more important role are discussed in a separate section.

2. Lexical diffusion

The overwhelming lexical regularity of sound change is an important phenomenon in historical linguistics both theoretically and because of the role it plays in internal and comparative reconstruction. A point that was frequently missed in the past is that a lexically regular outcome does not mean that while the change is ongoing it is also lexically regular, though Hugo Schuchardt made this point over a hundred years ago (Schuchardt 1885[1972]). The current discussion of lexical diffusion is not restricted to changes that have left some lexical items with the relevant phonetic conditioning unaffected, but rather to what is happening while the change is in progress. I argue that the pattern of lexical diffusion that is observed while a change is in progress provides important evidence about the mechanism that is driving the change.
Our understanding of how sound change affects the lexicon has been hampered in the past by the acceptance of the phonemic principle, or the premise that cognitive representations of sounds are abstractions over surface realizations and that words in memory are written out in a uniform abstract script that omits variation, predictable or otherwise. In such a conception of cognitive representation, sound change could be phonetically gradual, because the variants of a single phoneme could gradually mutate, but it would have to be lexically uniform – all instances of a phoneme must change in the same way at the same time. Or, a sound change could diffuse gradually through the lexicon, but it would have to be phonetically abrupt, changing phonemes, not phonetic variants. Phonemic theory explicitly rules out phonetic variation that is specific to particular lexical items, so there can be no change that creates gradual phonetic variation in particular words.

Against this theoretical background it is easy to understand the stand taken by Labov in his influential 1981 paper and in subsequent work. Labov distinguishes two types of sound change: the first he calls Neogrammarian sound change; the outcome is lexically regular and he characterizes it as ‘change in low-level output rules’; the second, he calls Lexical Diffusion change and he characterizes this type as ‘abstract phonological change’ or the replacement of one phoneme by another; this type diffuses gradually through the lexicon. Labov (1981) and (1994) reports that he finds little evidence of lexical diffusion in the many gradual vowel changes he has studied, though he does attribute the split of /æ/ in Philadelphia to lexical diffusion. He also notes that lexical diffusion is not to be expected for gradual consonant change in manner of articulation and some other consonant changes. Blevins (2004) and Phillips (2006) provide critiques of Labov’s theory. Blevins (2004: 266) recognizes the role of frequency of use in sound change which arises from variation, but she chooses not to label this as ‘lexical diffusion’. Phillips provides strong empirical support for the role of high frequency words, low frequency words and word classes in lexical diffusion. Her theory was influenced by and has provided additional evidence for the proposals in Hooper 1976.

Kiparsky (2003) expresses a view similar to Labov’s: he assumes that phonetically gradual change does not proceed by gradual diffusion through the lexicon. He proposes that only lexical rules (within his framework of Lexical Phonology) exhibit the effects of lexical diffusion. Moreover, Kiparsky argues that the mechanism for changes exhibiting gradual diffusion is analogy. This claim may well be correct for some of the ‘lexical diffusion’ changes that have attracted the most attention. Phillips (2006) reviews phonetically abrupt changes, many of which involve changes in stress pattern, and concludes that they diffuse gradually through the lexicon by analogy to high type-frequency patterns.1

1. See also Page (1999) for a proposal involving irregular changes motivated by prosody.
What the proposals by Labov and Kiparsky concerning lexical diffusion have missed is that in many cases phonetically gradual change also shows a gradual progression through the lexicon. This point is worth emphasizing because it is the nature of this progression that provides a window on the causes of sound change. The changes in question usually go to completion by affecting the entire lexicon. Thus many changes documented in the past with regular ‘before’ and ‘after’ stages might well have moved gradually through the lexicon. Therefore, in order to study the lexical effects of phonetically gradual changes, it is necessary to study change in progress. Studies that have been able to capture changes in progress (Hooper 1976, 1981; Phillips 1984, 2006; Oliveira 1991; Krishnamurti 1998; Bybee 2000, 2002a and b; Hansen 2001) demonstrate that there are many sound changes that are gradual both phonetically and lexically. From a synchronic perspective this means that words with the same ‘phonemic’ strings might have different, lexically-specific, ranges of variation. The most common pattern uncovered in such investigations is a frequency effect: variation in high frequency words shows that they undergo change earlier than low frequency words.

The following list documents lexical diffusion patterns that show high frequency words undergoing change earlier or to a greater extent than words of lower frequency. In most, perhaps all, cases the phonetic environment provides a stronger predictor of the change than token frequency as the changes are caused by phonetic factors, but in all of these studies, the phonetically gradual change occurs earlier and to a greater extent in high frequency words. The evidence is strongest for reductive changes, but there is also some evidence that vowel shifts and retiming changes can follow such a pattern.

a. Vowel reduction and deletion
   - Pre-stress vowel reduction in English (Fidelholtz 1975) and Dutch (Van Bergem 1995)
   - Reduction and deletion of schwa in American English (Hooper 1976; Patterson et al. 2003 as reanalyzed by Phillips 2006)
   - Reduction of vowels in hiatus in Spanish (Alba 2008)

b. Consonant reduction
   - t/d deletion in American English (Gregory et al. 1999; Bybee 2000, 2002b)
   - Final [t] deletion in Dutch (Goeman & van Reenen 1985; Phillips 2006)
   - Deletion of [ð] in Spanish (D’Introno & Sosa 1986; Bybee 2001, 2002a)
   - Reduction of Spanish [s] to [h] in syllable-initial position (Esther Brown 2004; Raymond & Brown in press)
Reduction of Spanish \([s]\) to \([h]\) to \(\emptyset\) (Earl Brown 2009)
fricativization of voiceless stops in English RP (Buizza & Plug 2010)
flapping in American English (Gregory et al. 1999; Patterson & Connine 2001)

– w-deletion in Danish (Pharao 2010)

c. Vowel shifts

– Diphthongization of Middle English \([i:\]\) and \([u:\]\) in Middle English dialects (Ogura 1987, 1995)

d. Assimilation and retiming or overlap

– Palatalization of \([tj]\) in American English (Bush 2001)

– Middle English preconsonantal diphthongization (Phillips 2006)

– Vowel changes in hiatus in Spanish (Alba 2008)

In all of these studies there is phonetic variation in the described phonetic environment, and the results show that higher frequency words have more of the innovative variants. The results are not due to just a few high frequency words skewing the pattern: some studies remove the highest frequency words and still find an effect; others use log frequencies which downgrade the effect of the highest frequency words (Gregory et al. 1999). Further study of changes that occur at word boundaries indicate that the rate of change of a given word is not just due to its frequency of use, but rather its frequency of use in the conditioning environment (Bybee 2002b; Brown & Raymond in press, Raymond & Brown in press).

Some of the phenomena listed above involve diachronic changes that occurred in the past and are considered sound changes (diphthongization of ME \([i:\]\) and \([u:\]\) [Ogura 1987], ME preconsonantal diphthongization [Phillips 2006]). In contrast, skeptics might view the ongoing vowel reduction in American English and Dutch and reduction of vowels in hiatus in Spanish (Alba 2008), or \(/t/\) and \(/d/\) deletion and palatalization in American English as only synchronic variation, not sound change. However, these represent changes that have occurred in languages as sound changes and in each case there are already indications of a permanent change in the language, if only in selected words. The other cases listed above – reduction of final and initial \(/s/\) in Spanish, deletion of \(/\emptyset/\) in Spanish, deletion of \(/w/\) in Danish, fricativization of voiceless stops in British RP – are all taken by the researchers to be sound changes in progress, with excellent evidence of the changes progressing across generations and across dialects. Indeed, it is commonly assumed that synchronic variation represents sound change in progress (Labov 1994; Hansen 2001; Guy 2005 inter alia).
3. The role of word frequency in sound change

As mentioned above, the reason for highlighting lexical diffusion in a theory of sound change is that it provides a window on the causes of sound change (Hooper 1976; Bybee 2001). The patterns of diffusion based on word frequency are compatible with certain mechanisms of change and not others. Change that diffuses from high frequency words or phrases to lower frequency ones is indicative of processes that occur in highly practiced behavior, i.e. the automation of neuromotor routines. This can be compared to a pattern in which low frequency words are affected first, notably changes of an analogical nature: analogical leveling affects low frequency words before high frequency words.

These two distinct patterns of lexical diffusion are due to two different processing mechanisms: the first is due to the domain-general process of automation of the production of repeated behaviors which progresses more rapidly with more repetition. As we will see below, this type of change occurs gradually and creates new gestural configurations and thus can create new segments. The second is due to the greater entrenchment of high frequency words with respect to their internal structure, both morphological and phonotactic. Entrenchment or lexical strength is built up through repetition and enhances accessibility (Bybee 1985, 2001, 2010). Easily accessible items are not likely to be remade on the basis of more general patterns of structure. This type of change occurs when a pattern with high type frequency serves as the model for a new formation. Thus this type of change, based as it is on existing patterns, will not create new segments. We will see in Sections 6.2 and 6.3 how this mechanism of change can apply in purely phonological cases.

Let us turn now to the first type of change – that which affects high frequency words and phrases first. The solid documentation of changes that are both phonetically gradual and lexically gradual has important theoretical consequences. As argued in Bybee (2000, 2001) and Pierrehumbert (2001), the gradual phonetic change in particular words is not predicted by theories in which only abstract phonemes exist in memory and all detail about how they are realized in context is lost. Rather, such facts require exemplar representation, where the cognitive representations of words consist of exemplars representing the phonetic variants of the word that the language user has experienced. Exemplar models also provide a natural way to represent change that occurs more rapidly in high frequency words, as a word that has undergone reduction in production will have an impact on memory representation, adding reduced exemplars or strengthening reduced exemplars that already exist. Frequency of use, then, does not cause a particular change to occur in a particular way. For changes due to automation of production, the phonetic environment is the ultimate cause, but the change progresses
faster in high frequency words because they are more often exposed to the production pressures that cause the change (Moonwomon 1992; Bybee 2000; Pierrehumbert 2001).

4. An articulatory basis for sound change

As shown above, the literature on sound changes that affect high frequency items first shows strong evidence that both reduction and assimilation are affected by frequency of use. From a gestural point of view, these findings correspond well to characterization of casual speech processes and sound changes by Browman and Goldstein (1990, 1991, 1992), Pagliuca & Mowrey (1987) and Mowrey & Pagliuca (1995). The former describe casual speech processes as: “due to two gradient modifications to gestural structure during the act of talking – (a) increase in overlap and (b) decrease in magnitude of gestures” (Browman & Goldstein 1992: 173)

In a comparable way, Mowrey & Pagliuca (1995) hypothesize that sound change is always in the direction of Substantive Reduction (reduction in the magnitude of gestures) or Temporal Reduction (by which gestures are compressed temporally and therefore overlap). Both of these proposals cover the two most common types of sound change – assimilation and lenition. The less common types – fortition and dissimilation – are mentioned briefly below.

The terms ‘automation of production’ or ‘automation of neuromotor routines’ are appropriate for several reasons. First, these are domain-general terms and therefore relate linguistic behavior to behavior in other domains. Second, it seems wrong to characterize reduction and assimilation as due to least effort, laziness or sloppiness. In other domains, the reduction and overlap of motor gestures that comes with practice is regarded as a high level of efficiency and at times precision, not sloppiness. Furthermore, automation occurs in much the same way among speakers of the same dialect. It is not as if each speaker is ‘sloppy’ in some idiosyncratic way. We recognize dialects because of the high degree of similarity in the way reduction and coarticulation take place within a dialect. Thus the efficiency that comes with automation is well described by Lindblom (1990) when he says that the speaker is subject to a general neuromotor principle that balances timing against the degree of displacement of physical movements in such a way as to make actions more economical. Thus coarticulation as well as reduction facilitate production (Lindblom 1990: 425). This theory, then, provides an explicit description of the link between the two processes that some have felt intuitively (Bauer 2008).

The phonetic gradualness of sound change is consistent with the preceding considerations as well as with the theory that language change takes place while
language is being used (Croft 2000; Kemmer & Barlow 2000; Bybee 2002b, 2010). In exemplar models, which are often coupled with the usage-based approach to language, there are no discrete categories; rather all categorization is stochastic as is all input and output. Thus the rather contradictory notion that phonetic change is gradual but phonemic change is abrupt, which is a natural outcome of phonemic theory or any theory in which abstract representations are posited, can be abandoned in favor of a theory in which all change can be gradual. As certain phonetic variants grow more frequent while others become less frequent, an exemplar category can shift gradually following the phonetic change as it is implemented. The detailed empirical study of Beddor (2009), which examines both the articulatory and perceptual consequences of the phonologization of vowel nasalization in English, uses terms such as ‘phonological grammars’ and shows that they may be different for different listeners, but the data presented there is also consistent with gradual change in the speaker/listener’s cognitive representation. Thus ‘phonologization’ would occur when a phonetic feature, such as vowel nasalization, has grown long enough and strong enough to be used as a major feature of word identification, and this is a process that can occur gradually.

The lexical diffusion facts cited above tell us that the articulatory routines for individual words can change at different rates, but that in the end, once the sound change is complete, there is likely to have been a general change in the articulatory routines for the language. What then is the relation between general patterns and those for specific words? In an exemplar model, general routines arise because individual words are stored with their ranges of pronunciations. General routines are built up from practice with many specific routines. As speakers use language, this interaction between the specific and general continues. While individual words have specific routines associated with them, their use activates the more general routines as well. As mentioned above, a certain amount of online reduction is expected in production, especially in certain contexts of low prominence or high priming. So when a word is produced with reduction, that reduced variant has an effect on the word’s exemplar cluster or cloud as well as on the exemplar cloud at the more general level of the articulatory routine. The effect on the general routine could explain why the change eventually spreads to lower frequency words (see Wade et al. 2010). In this view, then, articulatory routines (or motor commands) are also stochastic: they encompass a range of variation in cognitive representation as well as in production. Given that such routines are variable and represent ranges rather than static points, there is no reason to suppose that they cannot also change gradually.
5. The roles of articulation and perception

In a series of papers over several decades, John Ohala has presented a theory of sound change based on the listener’s tasks (Ohala 1981, 1989, 1993, 2003). Two mechanisms are proposed, leading to fundamentally different types of change: (i) hypo-correction, which accounts for an (assimilatory) change becoming emancipated from its conditioning environment, and (ii) hyper-correction, which accounts for dissimilation. Blevins (2004) adopts much of Ohala’s model, but emphasizes the role of first language acquisition in sound change, a point I address below. Other researchers have particularly focused on the first type of change, presenting somewhat more nuanced and more empirically-based versions of Ohala’s model (Beddor 2009; Harrington et al. 2008; Kleber et al. in press). The differences between Ohala’s view and others will be outlined in this section, with particular emphasis on the role lexical diffusion can play in helping us find the right interplay of factors.

5.1 Hypo-correction: Cause of change or reaction to articulatory change?

Ohala intends the mechanism of hypo-correction to apply to sound changes that are similar across languages and therefore can be assumed to have a basis in phonetics. As mentioned above, most sound changes fall into this group and it is largely coterminous with the type of phonetically gradual changes I have just described above. It seems clear that the source for sound changes of this type are the patterns of coarticulation that exist in synchronic language. Since changes in both articulation and perception occur in sound change, the question arises as to which of these two sides – that of the speaker or that of the listener – sparks the innovation.

Ohala’s proposal is that a change in the listener’s perception is the first step towards change. He notes that laboratory studies demonstrate that for some coarticulation patterns, listeners normalize the input in perception, correctly attributing some aspects of acoustic values to the context; thus when a vowel is nasalized preceding a nasal consonant, the nasality may be attributed to the consonant, not to the vowel. Given that there is noise and ambiguity in the acoustic signal, the proposal is that the listener might on some occasion fail to normalize, erroneously failing to attribute a feature to the context and considering it instead inherent to the segment being analyzed – in our example, attributing the nasality to the vowel rather than to the consonant. When this listener turns speaker, then production

---

2. See Bybee (2009) for a critique of Blevins’ theory.
might be changed, resulting in a weakened nasal consonant and a robustly nasalized vowel.

As listeners have a finely-tuned perceptual system that allows details of pronunciation and coarticulation to be very similar across members of a speech community, the next question to ask is under what circumstances such errors of perception would occur. Ohala offers two answers: “First, the listener may not have the experience to enable him to do such correction. Children in the process of acquiring the phonology of their language are in this position as are adult second-language learners” (Ohala 1993:247).

This is the less compelling of the two suggestions, as the sociolinguistic facts show that neither children nor second-language learners are in a position to initiate a change in the language at large. As Labov has demonstrated, the speakers that are propelling change forward are teenagers (Labov 1982). Studies of cases where the input to young children shows phonetic or phonological variability find that children acquiring their language exhibit the same phonological, stylistic, grammatical and lexical constraints that surrounding adults exhibit (Patterson 1992; Roberts 1997; Chevrot et al. 2000; Díaz-Campos 2004; Foulkes & Docherty 2006). These studies do not find evidence that children are the innovators. Of course, children may push forward certain patterns of change that are already in progress in their language, but in this regard their contribution is similar to that of adults.

Ohala gives the following as his second reason for failure of the listener to normalize the input (hypo-correction): “A second reason for hypo-correction is that a listener may, for various reasons, fail to perceive or to attend to the phonological environment which causes, or as phonologists usually put it, ‘conditions’ the variation” (Ohala 1993:247).

He goes on to point out that many assimilatory sound changes result in the loss of the conditioning environment, as when nasal consonants are reduced and lost leaving distinctly nasalized vowels. He further asserts that loss of the conditioning environment “is an important aspect of many hypo-correction sound changes...” (Ohala 1993:247).

This point leads directly to the idea that small changes in the strength of the ‘predictable’ feature (say vowel nasalization) and weakness in the conditioning environment (shortening of the nasal consonant) could lead to a change in how a listener parses these features. That is, to continue with our example of vowel nasalization, at a certain point, the nasality becomes strong enough and the nasal consonant weak enough that the listener attributes the nasality to the vowel or the whole syllable coda rather than to the consonant. This in turn can lead in production to a further strengthening of the vowel nasalization and weakening of the consonant. Under this interpretation, many small changes in both perception and articulation propel a change forward.
This view, however, differs from Ohala’s in the following way: a change in the coarticulatory pattern occurs first and the perceptual reinterpretation follows, whereas, in Ohala’s view, as I understand it, the proposal is that the perceptual change triggers the articulatory change. Below I will offer arguments as to how ‘articulation in the lead’ explains more of the data associated with sound change. However, first we must address the explanation for why articulation changes.

For any theory of sound change, postulation of the initial step is the most difficult as the earliest stages are the least accessible to investigation. As Ohala does, I start with the observation that coarticulation produces variable results. The variation is not all random; rather there are biases that make variation in certain directions more common than others. Such biases have been laid out in Lindblom’s theory, mentioned above. For instance, both reduction and overlap of gestures are a normal part of the automation of production. Other biases may have to do with prosodic and rhythmic structures, which give more prominence to some segments of the speech chain and less prominence to others. These biases apply in all speech events and have the effect of pushing variation in a consistent direction. All production pressures are highly dependent upon context, including degrees of redundancy and speaking rate. Language users walk a fine line between economical production and effective communication and as we know, economical production is more acceptable in casual and intimate social situations and in higher frequency lexical words and phrases. These are very likely the loci for the initiation of a sound change, as reduction and overlap may be allowed to proceed a little bit farther in certain situations and with certain words and expressions.

This explanation can be compared to that of misperception theory – that a random error in the parsing of features, with no articulatory trigger for this error, becomes a sound change. Such a theory requires that many speakers make the same ‘error’. The following are some of the reasons why the ‘articulation in the lead’ theory seems more plausible.

First, the speakers in the same speech community are subject to the same phonetic biases, which come from the nature of the physical apparatus and the nature of language-specific coarticulation and prosody. Thus a change does not have to start with an error or with just one person; rather, all (or most) speakers are inherently headed in the same direction. Extremely minimal changes take place within individual usage-events and language users track these minimal changes in their input and output long before they rise to the level of consciousness.

Second, because change takes place as speakers use the words and phrases of their language, change takes place within lexical items. As production is more compressed and reduced in intimate social situations and in high frequency words and phrases, this is where change begins, only to spread later to the whole lexicon. Thus an articulatory-based account for the initiation of sound change is consistent
with the lexical diffusion evidence presented above. There has been no account of how change spreads from one word to another in the perception-based theory. Indeed, changes motivated by misperception should be more likely in unfamiliar, infrequent words, rather than familiar, frequent ones (see Section 6.2).

Third, the articulatory theory can also account for exceptional changes such as those found in extremely high frequency phrases such as I’m going to to [amănă] or I don’t know to [aṟōno]. Such changed phrases are not irrelevant to the understanding of regular change because they involve extreme versions of more regular ongoing coarticulatory processes, such as vowel nasalization, consonant deletion and flapping.

Fourth, the hypo-correction hypothesis is illustrated with assimilations and the experimental studies of perceptual compensation involve assimilation processes, such as vowel nasalization, [s] vs. [ʃ] in vocalic context and fricatives after nasals (Mann & Repp 1980; Kawasaki 1986; Manuel 1995, among others). The same normalization scenario has not been shown to be applicable to reductive changes such as intervocalic spirantization or flapping of [t] and [d] in unstressed syllables. The difference between assimilation and reduction in terms of perceptual compensation and possible misparsing is that in assimilation, the eventual change often eliminates the conditioning environment as the feature is shifted to adjacent segment (Ohala 1993), while in reduction, the conditioning environment remains, but the affected segment can eventually delete as it further reduces. Thus in the case of intervocalic reduction of a stop to a fricative or flap, it has not been demonstrated that listeners assign the reduced variant to the context, nor can the reassignment of the fricative or flap from the vocalic context to the changed segment explain why such segments continue reducing. It appears that reduction of consonants is more likely the result of hypo-articulation as described by Lindblom.

Let us now consider some recent studies that have examined Ohala’s theory in the laboratory by measuring the variant segments undergoing change and testing the perception of contrasts in subjects. Beddor’s (2009) detailed study of vowel nasalization in American English finds, as other studies have, that there is covariation between the nasalization on the vowel and the duration of the nasal consonant. In addition, she finds that despite considerable variation across productions, the velum opening gesture is fairly stable. Her perceptual experiments showed that listeners were tuned to the overall nasalization across the syllable rather than to the details of the duration of the nasal consonant, or which segment the nasalization belonged to. She thus concludes that even accurate listeners can participate in a sound change. She also notes that the articulatory conditions under which nasal vowels are produced and nasal consonants weaken are the same as the conditions under which these changes proceed in sound change across languages. Thus her theory, based on her laboratory findings, is similar to the proposal made here,
giving no special role to perceptual errors, but proposing instead that production and perception change together.

Harrington et al. (2008) and Kleber et al. (in press) study the fronting of high back vowels in Standard British English. The first study shows that younger speakers, whose high back tense vowels were more fronted, also showed less compensation for context in the identification of high front vs. back vowels. This finding is consistent with the theory that giving up the normalization of the vowels in context leads to change. As the authors note, however, it is also consistent with other interpretations, in particular that the change is conditioned by consonantal environment, the vowel occurring more frequently in words with surrounding alveolars. As this is not the case with the high back lax vowel, the study by Kleber et al. sought to test the same hypotheses on production and perception data on this vowel. Again, it was found that the younger speakers had more fronted variants and that they compensated somewhat less than older speakers for consonantal context. Kleber et al. lean towards an interpretation consistent with Ohala's theory: that waning listener compensation for coarticulation is responsible for sound change. However, since the younger subjects already had vowels that were more fronted than the older subjects, we still do not know if the change in perception or production came first.

Harrington et al. (2011) provide a possible articulatory explanation for the tendency of high back vowels to front: such vowels, especially if they are extremely peripheral, involve a high articulatory cost, are prone to target undershoot, and have a propensity to encroach on the perceptual space of front vowels. Thus in this well-studied case, the facts are consistent with both the articulatory and perceptual accounts.

In trying to pull apart the question of whether the initial impetus for a sound change comes from production or perception, a major obstacle is the fact that perceptual similarity is usually paralleled by articulatory similarity. Thus it is not enough to say that x changes into y because they are perceptually similar if they are also articulatorily similar and vice versa. For this reason it is important and useful to look at other properties of a sound change, as mentioned in the introduction. For present purposes, patterns of lexical diffusion can be taken into account. As mentioned above, the pattern of change spreading from high frequency words to low is not the only pattern that has been observed (Phillips 1984, 2006). While the changes affecting high frequency words first are compatible with an articulatory source for sound change, a change that affects high frequency words last is not. Thus patterns of lexical diffusion, when they are available, could be quite valuable

---

3. The authors used type frequency despite the fact that they cite Pierrehumbert's 2001 model, which relates sound change to token frequency.
in helping to identify the source and cause of a particular sound change. Other properties of the change, such as those mentioned in the introduction, can also be referenced to help find an explanation, as Ohala (1993, 2003) has argued.

6. Candidates for perceptually-motivated change

The assumption that language change takes place in language use has been very useful in explaining the semantic, phonetic and morpho-syntactic changes that take place in a process such as grammaticalization. For this process we see that the cognitive processing mechanisms affecting both speaker and listener – such as categorization, inferencing and entrenchment – make important contributions to change. It follows that for phonological change, all the elements that go into online processing and storage are candidates for sources of change. In the preceding we have examined sound changes that are phonetically gradual, cross-linguistically common and related to coarticulation processes. These, I argued, come from articulatory innovations arising from the automation of production and are likely to affect high frequency words before low frequency words. However, if a change lacks one or more of these properties, then the source of the innovation may not be automation of production. In this section we examine a few such examples for which a perceptual source might be hypothesized.

6.1 Dissimilation

Ohala (1993, 2003) proposes hyper-correction to account for dissimilations. This seems to be a plausible account of dissimilation at a distance, though some of the examples attributed to this mechanism may have other explanations. Here the idea is that the listener knows that certain features spread over surrounding segments and corrects for that in interpretation. However, if two segments in a word have the same features inherently, correcting for that feature in one segment may cause it to be interpreted as lacking that feature. For instance, certain Latin words that had two /r/’s in consecutive syllables, were changed in Spanish such that the second /r/ became an /l/, because rhotic properties of the second /r/ were taken to be an extension of the first /r/: robur > roble ‘oak’, carcere > cárcel ‘jail’, marmore > marmol ‘marble’, arbore > árbol ‘tree’, Old Spanish (OSp) miécores > miécoles ‘Wednesday’. There are also cases where the first consonant changes presumably for the same reason: taratrum > taladro ‘drill’ and OSp cerebro > OSp celebro ‘cerebrum’ (examples from Menéndez-Pidal 1968: 182–3).

Two facts are consistent with this account: first, such changes are usually described as ‘sporadic’ (Menéndez-Pidal 1968: 181–182), meaning they are not
lexically regular. Thus they could be the result of a reinterpretation that occurs one word at a time. A hypothesis worth investigating would be that low frequency words are more prone to this type of change than high frequency words. Second, it is also significant that no new segments are created in such changes. This fact points to a type of reanalysis, as Ohala predicts, whereby listeners are sorting out the gestural configurations of a word and assigning them to known segments. In contrast, novel segments (nasalized vowels, palatalized consonants) are often created in other types of sound change.

6.2 Changes affecting low frequency words first

The hypothesis that different patterns of lexical diffusion point to different sources of change has been extensively examined for a range of types of change, both phonological and morphosyntactic (Hooper 1976; Phillips 1984, 2006; Bybee 2000, 2010). We have already discussed the connection between sound change and the spread from words of high token frequency to those of low token frequency. The opposite direction of spread – from low frequency to high frequency – is commonly found in changes that can be considered analogical; that is, changes by which the structure of a set of forms in the language affects a change in a form such that it will match this structure. Thus analogical leveling takes place earlier in low frequency forms and sometimes does not affect high frequency forms at all. The reason for this is that high frequency forms are strongly represented in memory and easy to access, so there is not reason to re-form them on productive models. Thus change affecting low frequency words first indicates that the form of such words presents a challenge to the listener or learner with the result that such words are remade on more familiar patterns. Productive patterns are those which apply to a large number of words in the language, i.e., those that have a high type frequency (Guillaume 1927 [1973], Bybee 1985, Hay & Baayen 2002).

As an example of a sound change that progressed from low to high frequency words, consider the change studied by Phillips (1984), the unrounding of Middle English /œ:/ and /œ/. These front rounded vowels lost lip rounding and merged with /e/ and /e/. Examples are deop > deep, beon > be, seon > see. This change is captured in progress in the text *Ormulum* from about 1200 A.D. The author was interested in spelling reform, and used two spellings for the reflexes of this Old English diphthong: eo and e, often representing the same word in two different ways. Phillips analyzed the spellings in this text and found that among nouns and verbs, the less frequent words had more innovative spellings, i.e. those that showed unrounding of the vowel, than the more frequent words did.

Such a change is a candidate for change via the failure to perceive and learn the distinction between rounded and unrounded front vowels. Note that on the
articulatory level, the change is the loss of a gesture and this could be its motivation. However, loss of a gesture would more likely take place in high frequency words first, so the pattern of lexical diffusion suggests another motivation. Phillips (2006) indicates that the front rounded vowels had a much lower type frequency than the unrounded ones. And, as Phillips (1984) points out, having a mid front rounded vowel without a high front rounded vowel is typologically odd. This fact perhaps means that the unrounded front vowels were easier to acquire and to access, and front rounded vowels may have been difficult to perceive, acquire and access, especially in low frequency words. Note that in this case no new segment types were created. An interesting twist in Phillips’ results is that the preposition *between*’ between’, the adverbs, *quickly* and newly and adjectives (except for numerals) display the opposite pattern: in this set high frequency words show more innovative spellings. A mixed pattern of lexical diffusion suggests a mix of mechanisms. High frequency words that are usually unstressed may undergo the change as a type of reduction – loss of lip rounding in unstressed position, while low frequency words undergo the change for a different reason.

6.3 Changes that cannot be articulatorily gradual

While it is rare, sound changes do occur which defy an account in terms of articulatory gradualness. These would be cases in which a gesture appears that is made by an articulator that had not previously been involved. This would not necessarily include cases of labial-velars becoming labials, as argued in Ohala (1993), as the labial involvement is already present. In such cases, the change is the reduction of the velar gesture and the strengthening of the labial one. (See discussion below of strengthening or fortition). Also the rare cases of labial palatalization, while they seem to involve the introduction of a new gesture, have been shown to occur only where the labial was already palatalized or preceded a palatal glide, as evidenced in Moldavian, Polish and Tswana (Bateman 2010). Similarly, the /f/ that developed from Old English /x/ was always preceded by a rounded vowel (Pagliuca & Mowrey 1987; Browman & Goldstein 1991). Such cases, then, have a source for articulatory gradualness. However, if one wanted to consider further the role of perceptual ‘confusion’, an investigation of the lexical diffusion properties of such a change might be instructive. Provided one could find such a change in progress, the prediction would be that if perceptual confusion were present, the change would occur earlier in low frequency words and only later in those of high frequency.

A case in which it is more challenging to find a source for what appears to be a new gesture is the change in British dialects from /θ/ to /f/ and in some dialects,
/θ/ to /v/. Here a lower lip gesture replaces a tongue tip gesture. Ohala (1993) says that this is due to perceptual confusion, citing experiments by Miller & Nicely (1955). This case may be similar to the preceding case in that perceptual difficulties coincide with low type frequency: the interdental fricatives in English are consonants with a low type frequency, occurring in fewer words than /f/ and /v/. Thus in cases where perceptual signals are not so robust, due to the typically low amplitude of non-sibilant fricatives, there may be a tendency to interpret the signals in terms of the pattern with the highest type frequency.4

While the replacement of /θ/ for /θ/ is spreading through the working class urban dialects of England and Scotland, the change may already have progressed beyond the stage of phonetic motivation and initial lexical diffusion to have become an emblem of social identification. Clark & Trousdale (2009) report that among the subjects they studied who were members of a band in west Fife, Scotland, social factors were the most significant determiners of use of /θ/ for the interdental. They also report, as do Stuart-Smith & Timmins (2006), a greater use of /θ/ word-finally than in other positions, a trend that is consistent with perceptual difficulty as the source of the change. Both studies mention lexical diffusion of this change, though the pattern points more towards word classes (such as proper names and ordinals) as factors than frequency of use. As for other phonetic factors, the data presented in Stuart-Smith & Timmins (2006: 177–8) indicates that the shift to /θ/ is much more common in words in which the interdental precedes an /r/ (74%) than in other words (44%). Whether this trend is related to articulation or perception is yet to be determined. Note again, however, that the result of the change is not a novel segment, but the substitution of one that was already in the speakers’ repertoire.

6.4 Conclusion

In this section three types of cases were discussed as candidates for change that is based not on articulation but on perception. Since such changes have been less studied as a group than changes with articulatory motivation and may constitute a broad range of types, I have tried to point out various factors that might be taken into account in future studies, particularly patterns of lexical diffusion in terms of token frequency and possible effects of type frequency of existing patterns.

4. See Hume (2004) for an explanation for cases of consonant metatheses based on perceptual indeterminacy and the tendency to interpret indeterminate sequences as instances of the pattern with the highest type frequency.
7. A typology of sound change and phonological change

The principal argument of this paper is that a variety of factors can be referenced for investigating the sources and causes of sound change. The nature of the diffusion through the lexicon has been emphasized here, but other factors are also important and have been mentioned here as well – the phonetic gradualness of the change, the similarity to changes in other languages, whether the change occurs just within words, or can occur across word boundaries, whether or not novel segments can be created and whether or not the conditioning environment is lost.

The following tentative table offers two types of changes in sounds. The first, most common type, I would like to call ‘sound change’ proper, as does Ohala (1993), as it is the more constrained type cross-linguistically. In my view this type has an articulatory source, but also involves gradual changes in perception, as outlined in Beddor (2009). The other (phonological change) subsumes various motivations for change, including perception, as discussed in Section 6. It is possible, however, that they all involve replacement of phonological properties (segments or sequences) that are of relatively low type frequency by properties that have a higher type frequency.

Table 1 is intended as a hypothesis for future investigations. The hypothesis is that the properties in each column cohere and that documented changes will draw properties from one column only.

Other factors could of course be added to this table, including the domain of application of the sound change, i.e. whether or not it applies across word boundaries, which is a property of sound change proper, at least at its beginning. Also one could consider the question of whether or not the conditioning environment is lost, which is a possibility for sound change proper but not so for phonological change. However, the factors discussed here are sufficient to illustrate that there are many possible diagnostics for the source of sound change.

Table 1. A typology of sound change based on six factors

<table>
<thead>
<tr>
<th></th>
<th>Sound change</th>
<th>Phonological change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetic path</td>
<td>Reduction &amp; retiming</td>
<td>Not restricted</td>
</tr>
<tr>
<td>Phonetically gradual or abrupt</td>
<td>Gradual</td>
<td>Abrupt</td>
</tr>
<tr>
<td>Lexical diffusion</td>
<td>From high to low frequency</td>
<td>From low to high frequency</td>
</tr>
<tr>
<td>Lexical regularity</td>
<td>Lexically regular</td>
<td>Lexical exceptions possible</td>
</tr>
<tr>
<td>Directionality across languages</td>
<td>Unidirectional</td>
<td>Differs across languages</td>
</tr>
<tr>
<td>Resulting segments</td>
<td>Novel segments possible</td>
<td>Existing segments only</td>
</tr>
</tbody>
</table>
8. Fortition

A consequence of the hypothesis that sound change proper is highly constrained and is caused by the automation of production is that sound change is predicted to consist mainly of gestural reduction and increased overlap. However, there is a class of changes that do not seem to fit the profile of sound change proper. These are changes – fortitions – in which articulatory gestures seem to be strengthened. A brief discussion of such changes and how they might be investigated follows.

The most important task is to clarify what a fortition might consist of. First, the characterization of lenition as the reduction in magnitude of gestures would mean that a fortition is the increase in the magnitude or duration of a gesture. This definition provides a consistent indicator of this common type of process and sets to rest questions of whether or not changes such as [t] > [ts] and [p] > to [pf] are lenitions. As these changes are the reduction in the magnitude of the closure gesture, they are considered lenitions (Pagliuca & Mowrey 1987; Honeybone 2001; Bauer 2008; Buizza & Plug 2010). The characterization of these changes as fortitions rests on applying perceptual criteria rather than articulatory criteria. If we apply articulatory criteria consistently, such changes are lenitions.

A second major issue concerns the distinction between the true strengthening of gestures and changes in timing. As has been pointed out (Pagliuca & Mowrey 1987; Browman & Goldstein 1991) excrescent consonants, inserted vowels and diphthongization appear to be types of retiming changes, rather than changes that add in new gestures or strengthen existing ones. Devoicing is sometimes regarded as fortition (Bauer 2008); however, within a gestural framework devoicing involves opening of the glottis and cessation of vibration, usually in anticipation of the word or syllable end. As such, no gesture is strengthened in devoicing.

A major class of changes that appear to strengthen articulatory gestures involve glides, especially palatal glides. For instance, the Latin palatal glide or semi-vowel was described as differing little from the vowel [i] and the labial-velar glide similarly resembled [u] (Kent 1945). By Late Latin, these semi-vowels had become more consonantal and this trend has continued in Romance such that palatal glides in many dialects (e.g. of Spanish) have become fricatives and even affricates (Face 2003). There are also other attestations of glide strengthening, such as that discussed above in which palatalized labials become palatal consonants (Bateman 2010) while the labial gesture reduces and is lost. The properties of such changes need to be examined carefully with an eye to establishing patterns of gradualness, lexical diffusion and domain of application. It might be that these properties will help us establish the motivation and mechanisms for fortition changes.
9. Conclusions

This chapter has presented arguments for the hypothesis that many sound changes are motivated by, and have their origins in, articulation. The typology proposed isolates a common type of sound change that is the result of the automation of production. This type of sound change is both phonetically and lexically gradual in its implementation, though the result is usually lexically regular. The typology does not rule out the possibility of effects of perception on sound change, it rather refers to a number of factors, especially lexical diffusion patterns, as diagnostics for particular causes and mechanisms of changes in sounds. The emphasis on lexical diffusion patterns in the discussion is justified because their diagnostic value has been seriously under-appreciated in the examination of the causes of sound change.

References


Patterns of lexical diffusion and articulatory motivation for sound change


Patterns of lexical diffusion and articulatory motivation for sound change


Language: Learning and processing ed. by Dagmar S. Divjak. & Stefan Th. Gries Berlin: Mouton de Gruyter.


