Sound change

Andrew Garrett
UC Berkeley

“[T]he causes of sound change are unknown.”
— Leonard Bloomfield, Language (1933:385)

1 Introduction

If new words and lexical usages are the most recognizable aspect of language change, the emergence of novel pronunciation patterns and the existence of distinct accents of the same language are also familiar to casual observers. Despite its ubiquity, there is no generally accepted definition of sound change. It certainly includes phonologization, whereby an automatic phonetic property evolves into a language-specific phonological one (Jakobson 1931, Hyman 1977, 2012). For some analysts, the phenomenon of sound change also encompasses phonological change with other origins, such as dialect contact, paradigm regularization, or structural simplification. Defining the object of study is important when different definitions lead to different views of its mechanisms and properties. Nonetheless, in what follows I take a pragmatic approach, highlighting some phenomena, questions, and further literature that seem especially interesting at present, and setting others aside.

Students entering the field in 2013 will find a different landscape from that of a generation ago. Inquiry into the causes and pathways of sound change has been transformed by the work of five scholars in particular. Three of them, Paul Kiparsky (1988, 1995), William Labov (e.g. 1981, 1994, 2001, 2010), and Björn Lindblom (Lindblom 1990, Lindblom et al. 1995), have proposed influential structural analyses of changes within their sound systems. Two others, Juliette Blevins (e.g. Blevins and Garrett 1998, 2004, Blevins 2004, 2006a) and John Ohala (e.g. 1974, 1981, 1983, 1993), have also emphasized the role of articulatory detail and perceptual processes in determining the course of sound changes. The last decade has seen valuable synoptic research, including books (Blevins 2004, Solé and Recasens 2012, Yu 2012b), review articles (Bermúdez-Otero 2007, Hansson 2008, Hualde 2011, Garrett and Johnson 2012), and dossiers of sound changes (Blevins 2004, 2008a, Kümmel 2007; still essential are Grammont 1939, Hock 1991). All this work offers a fine array of starting points for new research projects.

Despite such work and the new questions that come with it, the theories that defined the field still guide many debates. The classical theory of sound change, due to the Neogrammarians (Osthoff and Brugman 1878, Paul 1880, Sievers 1901) and adopted in American structuralist thinking

---

1 Many thanks to Claire Bowern, Beth Evans, and Larry Hyman for helpful advice and comments on a draft of this chapter.
Bloomfield 1933), had two main tenets. The first and more celebrated, the so-called ‘Neogrammarian hypothesis’, was that sound changes are exceptionless, gradual, and imperceptible while underway. The second was that changes that are not grounded in articulatory processes have an entirely different nature; dissimilation was the parade example. The Neogrammarians also contended that gradual articulatory change is caused by random variations in exemplar memory. Issues raised by these claims are still of interest. For example, Ohala’s distinction between perceptual ‘hypocorrection’ and ‘hypercorrection’ is one proposed solution to the dissimilation problem, while Labov’s (1981, 1994) study of lexically irregular sound change is a well-known approach to the Neogrammarian hypothesis. And some current attempts to model sound change (e.g. Pierrehumbert 2001, Wedel 2006, Garrett and Johnson 2012, Kirby 2012) use exemplar models of phonetic knowledge that formalize ideas of Paul (1880) or Kruszewski (1887:145-146), who wrote that “each articulation is governed by the unconscious memory of similar articulations already produced earlier.”

Equally influential is the foundational work on language change from a socially contextualized perspective. Weinreich et al. (1968) frame the field through several key problems. The three problems listed in (1) also serve to organize this chapter.

(1) Three problems raised by Weinreich et al. (1968)

a CONSTRAINTS: What changes are possible and what are their preconditions?
b EMBEDDING: How is a change embedded in its linguistic and social system?
c ACTUATION: Why does a possible change take place in one language, but not in another language with similar conditions?

After a review in §2 of prototypical sound changes and other changes that should perhaps not be treated as sound change, this chapter is organized as follows. The constraints problem and the actual mechanisms of sound change are discussed in §3, the embedding problem in §4, and the actuation problem in §5. A brief conclusion follows in §6.

2 What sound change is and isn’t

2.1 Phonologization

Prototypical sound changes involve phonologization, in which a direct link can be found between a change and some automatic phonetic process: a precursor. In the examples that follow, it should go without saying that calling changes phonologization does not explain how they happen.

A typical case is Old English umlaut, illustrated for two vowel qualities in (2).
Old English umlaut in consonant-stem nouns (Brunner 1965)

<table>
<thead>
<tr>
<th>SINGULAR</th>
<th>PLURAL</th>
<th>UMLAUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>lūs 'louse'</td>
<td>*lūsi</td>
<td>&gt; lysis 'lice'</td>
</tr>
<tr>
<td>mūs 'mouse'</td>
<td>*mūsi</td>
<td>&gt; myss 'mice'</td>
</tr>
<tr>
<td>fōt 'foot'</td>
<td>*fōti</td>
<td>&gt; *fōt &gt; føt 'feet'</td>
</tr>
<tr>
<td>gōs 'goose'</td>
<td>*gōsi</td>
<td>&gt; *gōs &gt; gēs 'geese'</td>
</tr>
</tbody>
</table>

Certain plurals originally had a nominative plural ending *-i, which disappeared after causing vowel fronting in a preceding syllable. Fronting was not restricted to plural nouns (it caused the alternation in modern *old, elder*); it is also notable that the fronted mid vowel was unrounded (e.g. *fōt 'feet*) earlier than the high vowel (e.g. *lyss 'lice*). Umlaut as seen in this example is cross-linguistically common. It seems obvious that umlaut is related to the well-established coarticulatory pattern whereby, in V<sub>1</sub>CV<sub>2</sub> sequences, a back V<sub>1</sub> is automatically somewhat fronted by a front V<sub>2</sub> (Öhman 1966). Such a coarticulatory process was phonologized as umlaut in Old English.

Tonogenesis is a second well-documented sound change pattern with a clear phonetic precursor. An example is given in (3) from the Austroasiatic language Khmu (a.k.a. Kammu; cf. Svantesson 1983).

<table>
<thead>
<tr>
<th>EASTERN KHMU</th>
<th>WESTERN KHMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>buc</td>
<td>pūc</td>
</tr>
<tr>
<td>puc</td>
<td>pūc</td>
</tr>
<tr>
<td>glaŋ</td>
<td>klāːŋ</td>
</tr>
<tr>
<td>klaŋ</td>
<td>klāːŋ</td>
</tr>
<tr>
<td>jaŋ</td>
<td>cāŋ</td>
</tr>
<tr>
<td>caŋ</td>
<td>cāŋ</td>
</tr>
<tr>
<td>raŋ</td>
<td>rāŋ</td>
</tr>
<tr>
<td><em>h</em>raŋ</td>
<td>rāŋ</td>
</tr>
</tbody>
</table>

In the eastern dialect, an original voicing contrast is preserved and surface pitch differences are allophonic. In one of the western dialects, the voicing contrast has been replaced by a tonal contrast. Low tone is found on syllables with a former voiced onset, a higher tone on syllables with a former voiceless onset. It is well known that lowered pitch is an automatic phonetic side-effect of obstruent voicing (Hombert et al. 1979, Kingston 2011); tonogenesis in cases like (3) is obviously a phonologization of this effect.

A third and less widely-known sound change pattern is the emergence of obstruents after high vowels, illustrated in (4) from the Tangkhulic (Tibeto-Burman) language Huishu.
Here an epenthetic k arose after high vowels, but not (as the last example shows) after non-high vowels. Similar changes are documented in several unrelated languages. Mortensen (2012:443) suggests that their precursor is a realization of “the final portions of high vowels ... as devoiced or fricative off-glides”, as in French. This phonetic pattern, driven by aerodynamic constraints, is also responsible for phonological high vowel devoicing in many languages.

2.2 Analogue changes in phonological form

In typical cases of phonologization like (2-4), a change can be described in purely phonological or phonetic terms, with no additional lexical or morphological conditions; within its environment the change is regular. But there are other pronunciation changes whose causes do not involve phonetic precursors. One class of examples includes contamination, as in (5). In such cases only one or a few lexical items are affected, rather than all words with a certain phonological context, and the change reflects the influence of related words. For these reasons, analysts usually treat contamination not as ‘sound change’ but as ‘analogical’ or lexical change.

(5) Two types of contamination

a Sequence contamination (Osthoff 1878)

Example: Latin October → dialectal Romance *Octember (Old French Octembre), under the influence of September, November

b Lexical polarization (Malkiel 1951)

Example: Romance *sinistru ‘left’ → *sinestru (Old Spanish siniestro), under the influence of *destru ‘right’ (diestro)

There are other cases with similar properties in which the classification is less clear. An example is what Moulton (1960, 1967) called ‘morphophonemic analogy’. In this type of change, a phonological alternation that expresses some morphological relationship is generalized to create a new segment. Yet the context in which the new segment arises is defined morphologically, not (as in ordinary sound change) in purely phonological terms. This is shown by the Swiss German dialect data in (6).
Morphophonemic analogy in Swiss German dialects (Enderlin 1911, Wanner 1941, Moulton 1960)

a  Morphological vowel fronting: /o/ → /ø/
   Plurals: /bog/ ‘bow’ → /bøg/, /of/ ‘oven’ → /øf/

b  Sound change: /o/ > /ɔ/ before coronal obstruents and /r/
   ii  No change: /gold/ ‘gold’, /ops/ ‘fruit’, /xɔxx/ ‘to cook’

c  Morphological vowel fronting and raising in Schaffhausen
   Plural: /bɔd/ ‘floor, ground’ → /bød/
   Diminutive: /lɔxx/ ‘hole’ → /løxxli/

d  Regularized morphological vowel fronting in Kesswil
   Plural: /bɔrɔr/ ‘borer’ → /bœrɔr/
   Diminutives: /xrɔt/ ‘toad’ → /xrœtl/, /flɔss/ ‘castle’ → /flœssl/

e  Morphologically isolated forms in Kesswil: /frɔʃʃ/ ‘frog’, /gɔtti/ ‘godfather’

The morphologically conditioned vowel fronting process in (6a) is dialectally widespread, and common in categories such as plurals and diminutives; itdescends from a much earlier umlaut sound change. The vowel lowering sound change in (6b) is found in multiple dialects and may be assumed to have a phonetic basis. This sound change gave rise to a phonological asymmetry in the Schaffhausen dialect, as shown in (6c): plurals and diminutives of words with lowered /ɔ/ have a higher vowel /ø/. In the Kesswil dialect, this asymmetry was regularized as shown in (6d): nouns with /ɔ/ form derivatives with the corresponding front vowel /œ/, which was not previously part of the phonological system. But crucially, the replacement of earlier */xrøtl/ by /xrœtl/ did not involve a general lowering of /ø/ before coronal obstruents. This is shown by the morphologically isolated forms in (6e); only derived forms underwent the change (Enderlin 1911:34). If morphologically isolated forms as in (6e) happened to be lacking, there would be no way to tell that the morphophonemic analogy was not a general lowering sound change.

Morphophonemic analogy is relevant because its analysis affects our notions of sound change. If it is sound change, then sound changes can be morphologically conditioned. But this kind of change involves the generalization of a morphological pattern, in this case vowel fronting, so it resembles contamination as in (5) and other strictly ‘analogical’ changes (e.g. the extension of a strong verb pattern to English dig, whose earlier past-tense form digged was replaced by dug in the 14th century). A key question is whether there is a purely phonetic precursor stage, for example in this case a stage where */xrøtl/ had a partly lowered vowel, or whether morphophonemic analogy always involves a categorical replacement.

Another case involving generalization of a morphological or lexical pattern is the regularization of stress in English nouns. Disyllabic nouns more often have initial stress than final stress, for
example, a pattern to which speakers are known to be sensitive. French disyllabic nouns are often
borrowed into English with final stress, as in *fatigue*, *grenade*, or *review*, but sooner or later their
stress tends to be regularized. Some results are shown in (7), with completed regularization cases
in (7a) and currently variable words in (7b).

(7)  English stress regularization in borrowed disyllabic nouns (Svensson 2004)

a  Historical regularization: *aspect*, *castle*, *manner*, *nation*, *palace*

b  Ongoing regularization (stress varying by dialect or within dialects): *chauffeur*, *debris*,
*garage*, *massage*, *paté*

The phenomenon is relevant here because the change is lexically irregular, and would contradict the
claim that sound changes are regular if it is counted as a sound change. But changes of this type
appear to be grounded not in a phonetic pattern but in a morphologically defined regularity about
English nouns. If the key Neogrammarian insight was to distinguish between (phonetically driven)
sound changes and (morphologically driven) analogical changes, then changes like those in (7) may
not bear on the properties of the former. Yet the patterns seen in stress leveling (Phillips 2006), like
those of morphophonemic analogy (Kiparsky 1968, Robinson 1976), play a notable role in modern
discussions of sound change.

The changes in (6d) and (7) are sensitive to aspects of phonological form and may even approach
regularity, but they also differ significantly from ordinary sound changes. Certain contact-induced
changes comprise another category whose origins may differ from phonologization. For example,
postvocalic voiceless stops are preaspirated in some Hebridean English varieties in Scotland (Shuken
1984). This reflects transfer from Scots Gaelic, which has a phonemic contrast between unaspirated
and aspirated stops. The latter are preaspirated in postvocalic positions; cf. *boc* /pʰɔk/ ‘male goat’
vs. *bog* /pɔk/ ‘soft’ (Nance and Stuart-Smith 2013). The Scots Gaelic pattern in turn is attributed to
Scandinavian contact (Hansson 2001, Helgason 2002). Preaspiration does arise via phonologization
(Clayton 2010), but the mechanism of change in Hebridean English and Scots Gaelic may have been
quite different if we assume, say, that Gaelic-dominant speakers transferred allophonic patterns from
Gaelic into English (cf. Lucas in this volume). In such a scenario there is no need to suppose that
the phonetic precursor of preaspiration existed in local English dialects prior to the transfer.

In retrospect, if we do not know much about the possible contact languages, it may be difficult
to decide whether a regular change resulted from phonologization of a universal phonetic pattern
or contact-induced transfer of a phonological pattern. For example, the Celtic, Germanic, and Italic
branches of Indo-European innovated initial stress. This might have originated internally, or it might
be due to transfer from indigenous languages of western and central Europe; we do not know.

In sum, words change phonological form in ways that are variously similar and dissimilar to
prototypical cases of phonologization. Some pathways of phonological change (e.g. morphophonemic
analogy) mimic phonologization, while others (e.g. English stress regularization) are unlike
prototypical sound changes. Different researchers may find it useful to take a narrower or broader
approach to the subject matter. Here, for clarity of focus, phonologization is taken as the prototype.
3 Constraints on sound change

If prototypical sound change is phonologization, it would ideally follow that the typologies of sound change and possible phonetic precursors correspond perfectly. This hypothesis serves as a restatement of the Neogrammarian position. Problems are then posed by any mismatch between sound change and phonetic patterns. Studying such mismatches is a strategy for addressing the constraints problem: what are the preconditions for sound change? The following overview describes two basic mismatch patterns (§3.1) and then examines enhancement (§3.2) and ‘minor’ sound change (§3.3).

3.1 Phonetic patterns and sound change: Basic mismatches

One class of mismatches consists of regular phonetic patterns that never lead to sound change. For example, in Seoul Korean, a contrast between lenis and aspirated stops (e.g. /p/ vs. /pʰ/) was formerly primarily cued by voice onset time; more recently, F0 expresses the same contrast (Silva 2006, Kang and Guion 2008). This is a classic case of phonologization akin to that in (3). Still, as Kirby (2012:229) writes, “in addition to VOT and f₀, spectral tilt and the amplitude of the release burst are relevant perceptual cues to the initial onset contrast”, but only f₀ is being phonologized. Tonogenesis is a common change, but not ‘tilting’ or ‘burst amplitude enhancement’. From a broader point of view, this is the question of why only some phonetic patterns seem to play a role in phonology.

A second class of mismatches consists of sound changes that do not correspond to phonetic precursors. A notable subclass involves those that allegedly lack any phonetic motivation. For example, Blust (2005) presents a dossier of ‘bizarre sound changes’ like the development of a Proto-Manus prenasalized trill *dr* to Drehet kʰ (e.g. *dra* > kʰa ‘blood’); the change of initial *dw* to Armenian erk is a similar Indo-European crux (Hock 1991:583-584, De Lisi 2013). One obvious difficulty with this line of reasoning is that a series of sound changes, each phonetically grounded, can yield an outcome or synchronic pattern that corresponds to no phonetic process; on Blust’s dossier see Goddard (2007) and Blevins (2007, 2008b). The error in such cases is to assume that the telescoped result of multiple changes arose through a single sound change; admittedly, it may not be clear when telescoping is plausible rather than a hypothesis of convenience.

A convincing case of telescoping is described by Stausland Johnsen (2012) in Standard East Norwegian, which has a synchronic pattern whereby /t d n s/ become apical postalveolar [t̠ d̠ n̠ s̠] when preceded by an apical alveolar /r/, which deletes. Examples are shown in (8a), together with examples in (8b) of a second process, triggered by /r/; the examples in (8c) are for comparison.

(8) Retroflexion in Standard East Norwegian (Stausland Johnsen 2012)

\[\begin{align*}
\text{a} & \quad /\text{tɔːr}/ \quad \text{‘take’ (PRES)} + /-\text{də}/ \quad \text{‘it’} \rightarrow [\text{tɔːdə]} \\
& \quad /\text{tɔːr}/ \quad \text{‘take’ (PRES)} + /-\text{n}/ \quad \text{‘him’} \rightarrow [\text{tɔːn}] \\
\text{b} & \quad /\text{stæːr}/ \quad \text{‘steal’ (IMP)} + /-\text{din}/ \quad \text{‘your’} \rightarrow [\text{stæːdin}] \\
& \quad /\text{stæːr}/ \quad \text{‘steal’ (IMP)} + /-\text{n}/ \quad \text{‘him’} \rightarrow [\text{stæːn}] 
\end{align*}\]
The process in (8b) has an expected assimilatory profile (/rd/ → [d], etc.), but the one in (8a) is not so well motivated. As discussed by Stausland Johnsen, the explanation is that apical /r/ originally triggered a change of laminal alveolar /t d n s/ to apical alveolar [t d n s], still preserved in other dialects. The apical alveolars and postalveolars then merged as postalveolars, yielding a telescoped process by which /rd/ → [d], etc. The broader point of such examples, as with some examples in §2.2, is that we cannot always infer sound changes directly from phonological patterns.

Another kind of case where sound changes do not correspond directly to phonetic precursors involves what Buckley (2009) calls ‘generalisation of sound changes’. For example, he argues that the Gallo-Romance /ka/ > /tʃa/ change (e.g. Romance *karu > French cher ‘dear’ with additional changes), surprising because it does not involve a front-vowel trigger for palatalization, had several steps. First, /a/ was allophonically fronted to [æ] in open syllables, with palatalization then occurring before [kæ] in a phonetically unsurprising way. Palatalization was then later extended to affect all instances of /k/ before tokens of the phonological category /a/, due to the phonological relationship between the [a] and [æ] allophones of /a/. If this analysis is correct, it highlights an interesting phenomenon but creates a practical problem for historical linguists, since it means that the original context of a sound change may not be directly inferrable from its effects.

### 3.2 Enhancement in sound change

Even in the clearest cases of phonologization, the outcome is not identical to the input before the change; in a V₁CV₂ context, a coarticulated V₁ is not the same as the output of phonologized umlaut. Therefore any account of sound change must reckon with enhancement of phonetic precursors. Three enhancement patterns are described by Garrett and Johnson (2012). One involves increasing the magnitude of some existing gesture, as in the umlaut fronting of [u] > [y]: the eventual [y] outcome is different from a back vowel with anticipatory coarticulation.

A related pattern involves temporal realignment of an existing gesture; typically, features or segments shift to perceptually more salient positions. For example, based on acoustic data from Cleveland-area speakers born from 1878 to 1977, Moreton and Thomas (2007) argue that Canadian Raising, whereby /ai/ > /ʌi/ before voiceless codas as in ice (vs. eyes), originated in a coda offglide effect. In their analysis, /ai/ → [ʌi] → [æe] before voiced codas, with the resulting [æi] : [æe] opposition then shifting to an [ʌi] : [æe] opposition. This final opposition is perceptually more salient. Similarly, several cases of consonant-vowel metathesis involve temporal realignment of a segment to a perceptually more salient position. For example, in the Norman French dialect of Le Havre (Maze 1903, Grammont 1909, Blevins and Garrett 1998), preconsonantal r and œr underwent linear-order inversions whose results conform to the generalization that r occurs before labial continuants and œr before other consonants. Representative output forms are shown in (9), with standard French equivalents showing the original order of rhotic and schwa.
Rhotic metathesis in Le Havre Norman French (Maze 1903, Grammont 1909)

a /bɔrbɛ/ ‘ewe, brebis’
   /bɔrdɛl/ ‘suspenders, bretelle’
   /fɛʁlyk/ ‘tuft, freluche’

b /ɛprɔvɛ/ ‘sparrow-hawk, épervier’
   /fʁɛm/ ‘closed, fermé’

c /ɡɔʁduʃɛ/ ‘gurgle, gargouiller’

d /ɛkʁɔʁjɛ/ ‘crayfish, écrevisse’

The examples in (9a-b) underwent metathesis while those in (9c-d) did not. As Blevins and Garrett (1998) note, F3 is lowered in the environment of rhotics and labial consonants, “an especially salient effect for fricatives, with their relatively long VC transitions.” The change thus has the effect of shifting r to a more conspicuous position.

In a third pattern, which Garrett and Johnson (2012) call auditory enhancement, a new gesture is introduced. The result is thus distinct from precursor phonetics in a way that is rather different from the magnitude and temporal-alignment enhancement profiles just described. A good example is the development of (noncontrastive) labial protrusion on palatalized sibilants. This is a well-known feature of English and other languages, and while it is not a universal feature of such sibilants it can arise even in cases of noncontrastive palatalization. For example, Northern Paiute has a process whereby the single underlying sibilant is palatalized after /i/. As described by Babel et al. (2012), at least in the Mono Lake dialect the underlying sibilant is an alveopalatal /ʃ/, as in [kacca] ‘wing’. Its palatalized allophone, as in [pʰiʃaʔju] ‘good’, is “characterized by higher laminal contact of the tongue surface and lip rounding”; see also Babel (2009). Several authors have suggested functional phonetic reasons for the labialization effect, for example, that it lowers the frequency of frication noise and thereby enhances the contrast with other sibilants (Toda et al. 2003, Toda 2009). A similar enhancement pattern must lie behind f labial protrusion in languages such as English, where f is historically also the result of palatalization (from k). Such changes are more than the enhancement of an existing gesture in phonologization.

3.3 ‘Minor’ sound change

Certain processes were traditionally excluded from precursor-based sound change, because linguists saw no precursor articulatory patterns that might generate them. These “cannot properly be put on a level with ordinary sound-change”, in the words of Bloomfield (1933:390). As Hoenigswald (1964:202) saw it, linguists distinguished “the so-called minor sound change processes ... by setting them off from other, presumably ‘major’ processes, and then by hinting mysteriously that the minor ones somehow represent borderline phenomena of a kind which suggests that there is more behind, and that their exploration will widen the theoretical basis for our existing treatment of change — giving us, perhaps, an entirely new framework for entirely new and deeper insights.” Minor sound change was said to be characteristically irregular and structure-preserving, unlike ordinary sound
changes that are regular and often yield new sound patterns or segments, but Hoenigswald points out that these putative distinctions are not so clearly established (cf. Hock 1985, 1987, Hume 2004).

The three most important ‘minor’ sound change types are dissimilation, non-local displacement (metathesis), and unconditioned saltatory change (e.g. apical > uvular r, English dialectal θ > f). These have indeed played a key role in inspiring a new approach to sound change, emphasizing the role of perceptual factors in addition to articulation. The core case, dissimilation, has been analyzed in several influential studies by Ohala (1981, 1993; cf. Kiparsky 1995, Blevins and Garrett 1998, Guion 1998, Blevins 2004, Barnes 2006). In Ohala’s analysis, sound change arises through errors in the process of perceptual correction. Listeners ordinarily factor out (correct) the effects of coarticulatory processes in parsing speech, but they may sometimes fail to do so. If a listener takes an automatic coarticulatory effect as phonologically real, hypocorrection (undercorrection) has occurred and a change like umlaut or tonogenesis may arise. According to Ohala, the converse may also happen: a listener may overcorrect by attributing a phonologically intended phonetic detail to context. For example, hearing a word with two aspirates in proximity, the listener may interpret the aspiration on one as an automatic effect of the other, and may then suppress it when speaking. Hypercorrection of this sort is Ohala’s explanation for dissimilation in cases like the Greek version of Grassmann’s Law (e.g. *hékb̥o > ekb̥ ‘I possess’).

Whether ‘minor’ sound change types originate through processes that differ from those active in other sound changes is a focus of current research. For example, saltatory changes may take place via an intermediate stage involving double or secondary articulation. Thus the English x > f change (e.g. in rough) involved an intermediate [xʷ], and there is evidence that the θ > f change involves an intermediate [θʷ] in English dialects and elsewhere (Garrett and Johnson 2012). Similarly, the shift of apical to dorsal r in French and other European languages may have involved a doubly-articulated apical-dorsal stage in some cases (Jespersen 1889, Morin 2013); in other cases dorsal r was borrowed directly from languages in which the sound change was complete. Other kinds of change that have been attributed to perceptual misparsing can also be reinterpreted; for example, Garrett and Johnson (2012) contend that some non-local displacement (metathesis) patterns that Blevins and Garrett (1998, 2004) explained as perceptual in origin have sources in motor planning.

Even dissimilation, seen as the parade example of misperception since Ohala’s work, sometimes has other causes. In Chahar Mongolian, for example, there is an aspirate dissimilation process that is very similar to Grassmann’s Law (Svantesson et al. 2005, Svantesson and Karlsson 2012). In this dialect, aspirate dissimilation can be described as deaspiration of stops immediately before a short vowel followed by Cʰ or s (i.e. any [spread glottis] segment). Examples appear in (10).


a  *sykʰe > /sux/ ‘axe’
    *tʰapin > /tʰEpj/ ‘fifty’
    *kʰari > /xtj/ ‘return’
b  \(^*k^h\)\textit{idh}t > /\textit{kt}^h\textit{at}/ 'China'  
\(^*t^h\textit{ik}^h\textit{in} > /\textit{t}^h\textit{ix}/ 'ear'  
\(^*t^h\)\textit{osun} > /\textit{t}^h\textit{os}/ 'fat'  
c  \(^*t^h\textit{ahalsun} > /\textit{t}^h\textit{aas}/ 'paper'  
\(^*k^h\textit{amt}^h\textit{u} > /\textit{x}^h\textit{amt}/ 'together'  

Words without deaspiration are shown in (10a), and words with deaspiration in (10b), while (10c) shows that at most a short vowel may intervene between the two aspirates; /\textit{x}/ is the unconditioned reflex of \(^*k^h\), while deaspirated \(^*k^h > k\) is unchanged. An important phonetic detail is that postvocalic aspirated stops are preaspirated: /\textit{VC}^h/ \(\rightarrow [V^hC]\). The hypercorrection model would interpret the change in (10) as follows: assuming for discussion that aspirate dissimilation followed some other changes, in words like \(*[k^h\textit{tat}]\) a listener attributed all the aspiration surrounding the first vowel to the second stop, reconstructing the word phonologically as /\textit{kt}^h\textit{at}/.

The Halh Mongolian dialect does not have phonological deaspiration; cf. e.g. /\textit{t}^h\textit{ix}/ 'ear', /\textit{t}^h\textit{os}/ 'fat'. But in the precise /\textit{Ch}^h\textit{VC}/ and /\textit{Ch}^h\textit{Vs}/ contexts where the Chahar dialect would show deaspiration, the Halh dialect has a related pattern: the first aspirate has a significantly shorter VOT than in /\textit{Ch}^h\textit{VC}/ contexts where no aspirate follows (Ramstedt 1902). In the speech of one Halh speaker, for example, Svantesson and Karlsson (2012) found an average VOT of 72 ms for the initial [\textit{t}^h] in [\textit{t}^h\textit{al}^h\textit{s}] 'steppe', as opposed to 50 ms in [\textit{t}^h\textit{a}^h\textit{t}^h\textit{x}] 'to pull' and 49 ms in [\textit{t}^h\textit{os}] 'fat'. This reduction of aspiration duration is obviously the precursor to complete deaspiration as in the Chahar dialect. But with aspiration still present on both of the original aspirates, it is not possible to interpret VOT reduction as an effect of perceptual overcorrection; the hypercorrection account requires the dissimilating feature to be lost. Motivations of gestural organization or motor planning instead seem relevant for the first stage of the change (as in Halh), with the eventual loss of aspiration similar to the loss of any reduced gesture (as in apocope or final stop loss). For general discussion of dissimilation and motor planning see Garrett and Johnson (2012).

4 Embedding sound change

The embedding problem involves questions about the relation between a change and its social and linguistic contexts. Here I highlight several aspects of linguistic context in particular. American structuralist discussions of sound change often emphasized how changes affect phonemic systems. For instance, the influential discussion of Hoenigswald (1960) distinguishes not only between merger (in which a formerly phonemic opposition ceases to be contrastive) and split but, importantly, between primary and secondary split. In a primary split, a conditioned sound change causes reassignment of some members of one phoneme class to another phoneme class. So, after the well-known Latin change by which \(s > r\) between vowels (Weiss 2010), as in *\textit{es}̃\textit{o} > \textit{er}̃\textit{o} 'I will be', certain words that formerly had instances of the phoneme /\textit{s}/ now had /\textit{r}/. Crucially, /\textit{r}/ was already a contrastive segment in the language. In this approach, pronunciation changes that produce new phones merely yield novel allophones of existing phonemes and so, by themselves, have no impact on phonemic
structure. That impact occurs only indirectly with subsequent changes and, in particular, secondary split. This term refers to a phonemic split that results ('secondarily') from another sound change. A classic example comes from Old English (OE), in which fricative voicing was noncontrastive: fricatives were voiced in medial (voiced) contexts and voiceless elsewhere. In Middle English (ME), several processes conspired to render voicing contrastive, among them the loss of final schwas, as in the dative singular OE /wiːfe/ [wiːve] > ME [wiːvə] > [wiːv] ‘wife’. After these changes, since the OE nominative singular /wiːf/ [wiːf] underwent no surface change, the OE phoneme /f/ had split into two phonemes /f/ and /v/. Crucial distinctions are that only secondary split yields a new contrast in the phonological system, while only primary split involves a change in pronunciation in the ‘splitting’ category. But the process by which fricatives became allophonically voiced in some Old English contexts, since it had no impact on the phonemic system as a whole or on phonemic categorization in individual words, was not a major object of analysis in this framework.

Today, discussion of linguistic embedding has largely shifted to debate over the preconditions of sound change. Nowadays, if the constraints problem highlights questions about the precursors of new pronunciations, the embedding problem focuses on structural contexts in which they may or may not emerge. Phonological, morphological, and lexical contexts, treated in §§4.1-4.3 respectively, are widely discussed from this point of view. To the extent that any of these contexts play a role in limiting the set of possible sound changes, or favoring some possible changes, then (contrary to traditional thinking) sound change is not fully determined by phonetic precursor patterns.

4.1 Phonological embedding

Phonological embedding is intensely debated in current research. Three questions can be highlighted. First, do principles of universal grammar block some logically possible sound changes, even if there are phonetic patterns that would otherwise generate them? This is a question about the existence of inviolable universal constraints, since principles of universal grammar by definition apply in every language. As Greenberg (1966:510) put it, “no diachronic change gives rise to a synchronically nonexistent type.” In this context Kiparsky (2006) argues that a word-final obstruent voicing process is phonologically impossible, excluded by principles of universal grammar, and that even an otherwise possible sound change could not yield such a process; he outlines several such logical pathways, and contends that they are never found. In contrast Yu (2004) and Blevins (2006a,b) contend that word-final voicing, while rare, is attested and therefore not excluded by universal constraints. This case and the broader question require more study, and a larger dossier of relevant examples, for a resolution to emerge.

Second, relatedly, does an analytic or learning bias favor the emergence of certain phonological patterns in preference to others, given phonetic precursors that seem equally robust? Research by some investigators finds evidence for such bias, usually in artificial-learning experiments (e.g. Healy and Levitt 1980, Wilson 2003, 2006, Moreton 2008, 2010, 2012), but others disagree (e.g. Pycha et al. 2003, Seidl and Buckley 2005, Yu 2011). For instance, Seidl and Buckley (2005) studied the sensitivity of nine-month-old infants to various phonological patterns, some of which are phonetically
grounded and some of which are not. An example of the former in their study was a pattern in which labial consonants are followed by round vowels and coronal consonants are followed by front vowels; an example of the latter was a pattern in which labial consonants are followed by high vowels and coronal consonants by mid vowels. In head-turn preference experiments, they found that infants are “able to learn the rule in both the phonetically grounded and arbitrary conditions” (p. 309). The literature in this area has been reviewed by Moreton and Pater (2012), who conclude that “the broader picture is essentially that proposed by Bach and Harms (1972): Inductive bias, a property of the learner’s pattern-detection processes, facilitates faithful acquisition of simple patterns and rejection or innovative simplification of more complex ones, but is (relatively) insensitive to their phonetic motivation” (pp. 711-712). In other words, while not all research has distinguished between structural simplicity and phonetic grounding as factors that might promote phonological learning, when this distinction is made it is simplicity and complexity that play the major role.

Third, does a language’s system of phonological contrasts make certain changes relatively likely? A positive answer goes back at least to Martinet (1951, 1952, 1955, 2005), who identified structural influences of several kinds still investigated today; compare the discussion of enhancement in §3.2 above. For example, it may be that certain sound changes are possible only in languages with certain structural properties. This claim was made by De Chene and Anderson (1979), according to whom compensatory vowel lengthening only occurs in languages with a pre-existing vowel length contrast, but see Gess (1998) and Kavitskaya (2002). Similarly, Matisoff (1973) contends that voicing contrasts yield tonal contrasts, as in (3) above, only in languages with pre-existing tonal systems or through contact with tonal languages (cf. Svantesson 1983, Kiparsky 1995, Bermúdez-Otero and Hogg 2003). This claim is called into question by the examples of Korean (§3.1 above) and some New Caledonian languages (Haudricourt 1968, Rivierre 1993), in which aspirated-unaspirated stop contrasts yield tonogenesis, but it remains to be determined why, if Matisoff’s claim is incorrect, such changes are not more common in languages with contrastive voicing or aspiration.

The approach pioneered by Martinet is best known for dispersion-based analyses of sound change, especially chain shifts. For example, Martinet describes vocalic chain shifts in the Franco-Provençal dialect of Hauteville (ɛ > Æ > æ > a > õ) and the Azorean Portuguese dialect of São Miguel (a > õ > o > u > y). In each chain the end-point (õ in Hauteville, y in São Miguel) was previously absent from the phonological system; Martinet views all the steps as driven by contrast preservation. Another well-known dossier of vowel shifts putatively driven by contrast preservation is that of Labov (1994, 2001, 2010), who analyzes the Canadian shift (ɛ > æ > a) and Pittsburgh ñ lowering, both triggered by the cot-caught merger, and American English û fronting, among other North American English dialect changes (cf. Labov et al. 2006). Boersma and Hamann (2008) present a theoretically precise analysis of Slavic sibilant shifts in terms of auditory dispersion. Overall, the range of cases described by various linguists from various points of view is impressive and seems compelling, but since the examples are assembled anecdotally it would also be helpful to have evidence that they are not due to selection bias.
4.2 Morphological embedding

While the Neogrammian manifesto (Osthoff and Brugman 1878) is often cited for its claims about sound change, the Neogrammarians were as interested in morphological change (Morpurgo Davies 1978), and the manifesto itself was the foreword of the first issue of a journal called *Morphologische Untersuchungen*. Their great practical legacy to historical linguists is the distinction between regular sound change and regularizing analogical change as independent diachronic processes. Therefore the existence of analogical effects in the early stages of a sound change, prior to any phonemic split, has puzzled historical linguists. It is sometimes even denied: “Analogue extension or levelling of some phonological rule can take place only if that rule is no longer present in the synchronic grammar of the language” (Hogg 1979:58). But the existence of subphonemic analogy was clearly described by Bloomfield (1933:366): “When we observe sub-phonemic variants, we sometimes find them … systematized among forms, quite in the manner … of analogic change.” His example was a “Central-Western” American English shortening of /a/ before /ʌk, ʊp/ (e.g. dark, sharp) and before /ʌd, ðt/ followed by syllabic /ŋ/ (e.g. barter, Carter, garden, marten, Martin). But when /ʌ/ or /ŋ/ is what he calls a secondary suffix (e.g. agentive -er or deadjectival verbalizing -en), Bloomfield observes that forms like carter, harden, starter have the unshortened vowel due to the influence of unsuffixed forms like cart, hard, start.

Subphonemic analogical effects have been noted in connection with several other phonological processes, including the representative English examples in (11).

(11) Subphonemic analogy: Additional English examples

a. Mid-Atlantic English /æ/ tensing: the tense vowel of pad, etc. extended to padding, etc. (Trager 1940; cf. Ferguson 1972, Labov 1981, 1994, Kiparsky 1995)

b. Southern American English diphthongization: diphthongs in dare, steer, etc. extended to verbal dɛjɛrɪŋ, stɛjɛrɪŋ ~ stɛjɛrɪŋ, etc. but not to adjectival or nominal dɛnɪŋ, stɛnɪŋ, etc. (C.-J. N. Bailey, quoted by Wells 1982: vol. 3, p. 547)

c. Scottish Standard English /aɪ/: the lengthened [aɪ] of sigh, tie extended to sighed, tied, while side, side retain [aɪ] (Scobbie et al. 1999)

d. American English flapping: the [t] of negative, primitive, relative, etc. extended to nequtɪvɪsɪstɪv, prɪmitɪvɪstɪv, relatɪvɪstɪv, etc. (Steriade 2000; cf. Davis 2005)

e. Irish English prerhotic coronal dentalization (e.g. ladder, matter): the alveolar stop of loud, shout, etc. extended to louder, shouter, etc. (Downing et al. 2005)

The problem is that if influences from related forms play a role in the early stages of sound change, it may be hard to find empirical grounds to exclude morphology from the conditioning of sound change. Yet this traditional exclusion has been very productive in historical linguistics. Bloomfield’s example and the cases in (11) can be analyzed as examples of ‘quasi-contrast’ or ‘quasi-phonemes’ (Ladd 2006, Scobbie and Stuart-Smith 2008, Kiparsky this volume), a notion that may be useful in other settings. For example, in the earliest (Mycenaean) dialect of Greek, a change of intervocalic *s
> h was subject to certain ‘analogical’ exceptions even though it did not yield a phonemic split; the regular outcome is found in words like *skelesa > skeleha ‘legs’, but dative plural forms like tri-si ‘three’ and kh alkeu-si ‘coppersmiths’ (Bartoněk 2003) show the case ending -si, preserved through the influence of words where s was not intervocalic (e.g. dat. pl. i rík-si ‘hairs’). This case can be analyzed by treating [s] and [h] as quasi-phonemes in Mycenaean Greek, but we do not know why some allophonic distributions become quasi-contrastive while others do not. Future research may show in just what situations subphonemic analogy occurs.

4.3 Lexical embedding

There are two main ways in which lexical patterns are sometimes thought to influence or inhibit sound change. The first involves lexical frequency, a debated matter that raises the following puzzle. Many well-studied completed sound changes seem to show no effects of lexical frequency; all words satisfying the phonological conditions for the change undergo it. Unconditioned mergers present the most obvious cases. For example, when *kw > p in ancient Greek, no residual frequent words with *kw were left behind; English x > f in words like cough, and the standard dialect has no residue of frequent words with x; Proto-Polynesian *t > k in Hawaiian, and while there is a dialect where this did not occur, in the dialect of Hawai‘i all words were affected. The loss of final stops in ancient Greek was without exception, as were the loss of final consonants in most Oceanic languages and of final schwa in Middle English. On the other hand, it is well established experimentally that less frequent words tend to be pronounced with greater duration or articulatory effort (Guion 1995, Gahl 2008, Bell et al. 2009). Moreover, some leniting sound changes seem to show frequency-based lexical conditioning (Phillips 1984, 2006, Bybee 2001, 2002), though examples have not always been sifted critically; some involve lexical or analogical processes other than sound change (such as the generalization of stress regularities), while others are based on impressionistic classification of gradient phonetic processes. But the evidence as a whole cannot be dismissed, even if most lenition does not seem lexically irregular in retrospect and many examples of ongoing sound change studied by Labov (1981, 1994) were lexically exceptionless. The evidence points in conflicting directions, so more research is still needed.

Apart from leniting changes where frequency may be relevant, apparent lexically irregular sound changes (‘lexical diffusion’) can generally be understood as the result of dialect borrowing (contact between changed and unchanged varieties) and similar sociolinguistic processes. Examples are well studied in English dialects (e.g. Labov 1981, 1994, Harris 1985). A typical case in East Anglia is described by Trudgill and Foxcroft (1978). They discuss a merger between the reflexes of Middle English /ɔ:/ and /ʌ/, traditionally realized as East Anglian /u/ and /ʌ/ in words like go, road, moan, rose, sole and flow, know, low, mown, row, soul respectively. The two phonological categories have merged completely in London speech, but at the time of their fieldwork the contrast was maintained intact in Norfolk, the East Anglian county farthest from London; closer to London there was increasing evidence for merger. Trudgill and Foxcroft also compared their findings with dialect data collected in the 1930s and 1950s. Throughout the region, the merger was in favor of /ʌu/, with
words like *rose* and *sole* shifting. This shift appears to have been lexically irregular at any point in time, but overall the historical data show the merger diffusing from London, with the most frequent /ɔː/ > /uə/ word resisting longer than any other word: in the data from the 1930s and 1950s there were dialects, increasingly farther from London, in which /uə/ remained only in *go*. This apparent frequency-conditioned sound change was a case of dialect borrowing or sociolinguistic diffusion in which, just as in analogical change, the most frequent word resists replacement.

More generally, regarding lexically irregular sound change, the conclusions of researchers in two very different paradigms agree entirely. Referring to lexical exceptions to sound change, Bloomfield (1933:360) wrote that “factors of linguistic change other than sound-change will appear in the residual forms after we have ruled out the correlations that result from sound-change.” He meant by this that the regularity assumption brings to light forms that are due to analogy or dialect borrowing. Similarly, Labov (1994:542) concludes that lexical diffusion “is most characteristic of the late stages of an internal change that has been differentiated by lexical and grammatical conditioning, or has developed a high degree of social awareness or of borrowings from other systems . . .” His point is like Bloomfield’s: sound changes acquire lexical exceptions due to analogical processes, and borrowing or sociolinguistic processes.

A second way in which lexical patterns have been considered relevant for sound change involves the notion of ‘functional load’. A given contrast is said to have a higher functional load than another contrast if it is crucial in distinguishing more word pairs. From a set of anecdotally assembled cases, Martinet (1952, 1955, 2005) argued that a merger is less likely if the contrast being neutralized has a high functional load, and more likely if it has a low functional load. Martinet’s claim was challenged by King (1967) on the basis of a small corpus of Germanic sound changes, and has mostly been ignored in recent decades. Recently, however, the idea that low functional load favors neutralizing sound change has been supported in a quantitative study by Bouchard-Côté et al. (2013), based on a corpus of inferred sound changes in hundreds of Austronesian languages. Some historical linguists may be skeptical because the methodology for inferring the changes themselves is novel, but it will revive an interesting element of the embedding problem.

5 The actuation of sound change

There is a consensus that sound change emerges from phonetic variation and a process of selection from among variants (Ohala 1989, Kiparsky 1995, Lindblom et al. 1995, Garrett and Johnson 2012). But there is no general agreement regarding the selection process; the roles of perceptual processes (§3.3) and phonological and other contexts in (§4) in determining the character of sound change continue to be debated. The initiation of sound change also raises the broader actuation problem of Weinreich et al. (1968). Why does a change occur at a certain time in a certain language, but not in other languages or at other times with similar conditions? Some linguists have judged such questions unanswerable, because the crucial data seemed inaccessible. “The process of linguistic change has never been directly observed”, wrote Bloomfield (1933:347), adding that “such observation, with our present facilities, is inconceivable.” But tools have changed in the last eighty years, with current
research suggesting two approaches to the actuation question. One is essentially psychological; the other is social.

The psychological approach seeks to identify kinds of individual who differ from local population norms in ways that may be relevant for language change. Such individuals are presumably present in all populations, but in low numbers, and are implicated in the initiation of at least some changes. One possibility is that there are individuals who have distinctive or idiosyncratic articulatory patterns. In the production of /VN/ = [VN] sequences, for example, the duration of vowel nasalization is generally shorter if the nasal consonant is longer. Beddor (2009) finds that individuals differ in the predictability of covariation in their speech, and suggests that the common /VN/ > /V/ sound change originates when the speech of individuals with relatively more predictable covariation is perceived as evidence that vowel nasalization is as good a cue as the historical nasal consonant. Similarly, in an analysis of s-retraction in American English, Baker et al. (2011) show that even individuals who do not show s-retraction differ significantly in the degree to which they produce its articulatory precursor. They argue that interspeaker articulatory variability may favor the innovation of a sound change if some individuals’ articulations are so distinctive that they are perceived by listeners as a distinct target. Especially with the increased use of ultrasound data in phonetic research, we can expect more and more study of individual articulatory differences, some of which surely play a role in initiating sound change.

A second possibility is that there are individuals who have idiosyncratic perceptual patterns. For example, Yu (2010, 2012a) finds that individuals differ in the degree to which they compensate perceptually for coarticulation. Since perceptual compensation is a key factor inhibiting otherwise natural sound changes, individuals who compensate less may be likelier to take coarticulated speech or phonetic outliers at face value. Such individuals would at least in some cases then be the innovators of sound change.

The social approach to the actuation problem takes seriously the commitment of Weinreich et al. (1968) to language as a social object. From this perspective, a language change only occurs when a speech idiosyncrasy of one or more individuals acquires social value in a speech community. The consistent presence in every community of individuals with articulatory or perceptual idiosyncrasies would not address the actuation problem, since it would not explain why some communities develop a change with a certain social value while others do not. A sociohistorical explanation is thus also crucial. For example, Herold (1997) suggests that in one part of eastern Pennsylvania the cot-caught merger was triggered by massive European immigration, though in her otherwise compelling analysis the specific relation between the cause and the effect remains unclear. Similarly, Labov (2007, 2010) argues that the Northern Cities Shift was triggered by koinéization as a massive influx of Erie Canal workers representing different /æ/-tensing dialects interacted in northern New York State. Adults accommodate to interlocutors in speech interaction (Nielsen 2011, Babel 2011, 2012); Labov’s account depends on the idea that certain complex processes are hard for adults to learn in dialect contact settings. The unconditioned Northern Cities /æ/-tensing pattern is then a simplification of the more elaborate patterns of the older dialects of the eastern seaboard. The basic idea is related
to a point made by Milroy (1987:187-188): “One important corollary to the link between language maintenance and a close-knit territorially-based network structure is that linguistic change will be associated with a break-up of such a structure.”

In the end the two approaches must to be linked, since language is learned and known by individuals but used in social interaction. Every change must originate with individuals, but if it is to be observed it must diffuse in a speech community. Unlike in any previous decade, in this decade it seems possible to make significant progress linking the two approaches and thereby beginning to solve the actuation problem.

6 Conclusion

Bloomfield’s (1933) pessimism about understanding the causes of sound change reminds us that we have much to learn, especially in addressing the actuation and embedding problems of Weinreich et al. (1968). But much has been learned in the 80 years since Bloomfield expressed that pessimism. Students entering the field today are in a position to make contributions that will advance our understanding of sound change and its intricate connections with anatomy and physiology, general cognition, phonological and morphological structure, and lexical patterns. In the coming generation, we can expect their work to change the field as dramatically as it has been changed in the last generation.
References

Bouchard-Côté, Alexandre, David Hall, Thomas L. Griffiths, and Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. Proceedings of the National Academy of


De Lisi, Jessica. 2013. Feature metathesis and the change of PIE *dʒu to Classical Armenian -ɾk-. Diachronica 30.


Kruszewski, N. 1887. Prinzipien der Sprachentwicklung [Parts III-VI]. *Internationale Zeitschrift für allge-


———. 1967. Types of phonemic change. In *To honor Roman Jakobson: Essays on the occasion of his


———. 2006. Learning phonology with substantive bias: An experimental and computational study of velar


