Diachronic explanations of sound patterns

Gunnar Ólafur Hansson
University of British Columbia

— DRAFT VERSION, AUGUST 30, 2007 —

ABSTRACT: Phonological systems show clear signs of being shaped by phonetics. Sound patterns are overwhelmingly phonetically ‘natural’, in that they reflect the influence of physical constraints on speech production and perception, and categorical phonological processes often mirror low-level gradient phonetic effects. The question of how best to explain and model the influence of phonetics on phonology has been approached in different ways, one of which situates the locus of explanation in the diachronic domain of language change, in particular sound change. On this view, recurrent sound patterns merely reflect recurrent sound changes with phonetic origins, typically in speech perception. Explicit models of sound change are reviewed and illustrated: Ohala’s listener-based model and Blevins’ Evolutionary Phonology framework, and the relevance of exemplar-based models of production and perception is also noted. Current issues of controversy regarding the adequacy of diachronic vs. synchronic explanations for the typology of sound patterns are surveyed.

1. Introduction: the phonetic bases of phonology

As even a cursory acquaintance with phonology will reveal, the vast majority of sound patterns and phonological processes in the world’s languages show clear indications of being shaped by the physical constraints of speech production and perception. Much of phonology is thus arguably ‘natural’ from a phonetic perspective. Distributional patterns and asymmetries in the ability of different positions to support phonemic contrasts (syllable onset vs. coda, stressed vs. unstressed syllable) are strongly correlated with the relative salience of acoustic-perceptual cues and/or the relative magnitude of articulatory gestures in the positions in question. Common phonological processes usually have clear parallels in low-level patterns of phonetic variation, both in the articulatory domain (coarticulation, gestural undershoot, inter-gestural timing and overlap, boundary strengthening) and in the acoustic-perceptual domain (confusion, misperception; aerodynamic effects on the acoustic signal). Such obvious and pervasive correlations raise a number of ontological questions which cut to the conceptual core of phonology as a discipline. What is the nature of the connection between phonetic ‘substance’ and the higher-order patterning that phonology deals with? How close is this connection, and how—if at all—should we allow it to inform our theoretical models of the implicit knowledge speakers have of the sound patterns of their native language? To what extent, and in what manner, should phonetics be called upon to explain empirical generalizations about the typology of sound patterns and phonological systems? What bearing does all of this have on the question of
whether there exists a phonological module of Universal Grammar (UG) and what elements of innate knowledge such a module might consist of?

The notion that the typology of phonological well-formedness is shaped by phonetic content, mediated by some substantive notion of ‘markedness’, was an uneasy fit in classical models of generative phonology (cf. the famous Chapter 9 of Chomsky & Halle 1968). In the 1980s and early 1990s, phonetic content became more closely reflected through elaborate three-dimensional representations partly mirroring the structure of the vocal tract (feature geometry). Attempts were made to encode markedness in terms of formal complexity at the level of segments (radical underspecification, privative features) and of phonological rules (spreading and delinking vs. more substantial transformations); see Kenstowicz (1994) for a review. These developments went hand in hand with the appearance of explicit generative models of the phonetics-phonology interface (Keating 1990), an increased interest among phonologists in interface issues (Archangeli & Pulleyblank 1994), and the emergence of ‘laboratory phonology’ as a research field in its own right with a dedicated biannual conference series from 1987 onwards (Kingston & Beckman 1990). With the advent of Optimality Theory (Prince & Smolensky 2004 [1993]; see McCarthy 2007), well-formedness constraints on surface representations have become the ‘prime movers’ defining sound patterns, and the study of typological variation—modelled as permutations of ranked constraints—has taken on a far more central role. This has opened up the possibility of allowing synchronic grammars to incorporate functional explanations in terms of phonetic factors (cue availability, perceptual distinctness, articulatory effort) in a far more direct and explicit way than earlier frameworks permitted (see, e.g., Steriade 1999, 2001, Hayes 1999, Hayes & Steriade 2004, and many contributions in Hayes et al. 2004; cf. also Boersma 1998).

The explicit functional orientation of a great deal (though by no means all) of the work carried out under the Optimality Theory rubric in the last decade or more has reawakened an old line of criticism (Anderson 1981; cf. Sampson 1975), that grammar-internal explanations are often redundant in that they ignore the availability of an alternative explanation—often independently motivated, and functionally well understood—in the diachronic dimension of language change. According to this view, if we know (or can infer) how a particular synchronic pattern came into existence as the end product of familiar kinds of diachronic events (sound changes), and if we have a sound understanding of the nature and causes of such events in the ‘real-world’ conditions under which speech is produced and perceived by language users and language learners, then this is all that is needed. Consequently, we should not attempt to invent a synchronic explanation for the same set of facts, couched in grammar-internal terms, especially if this requires positing new theoretical constructs which are then attributed to the innate endowment of humans. In short, functional-phonetic explanations are to be sought and defined in the diachronic domain, not attributed to the inherent ‘design properties’ of the synchronic grammars internalized by speakers. The view that
Diachronic explanations for phonological patterns are not only a powerful tool, but a preferable way of deriving both the characteristic phonetic ‘naturalness’ and the occasional ‘unnaturalness’ of certain sound patterns, is the focus of this review.

The claim that explanations for synchronic sound patterns should be sought in their diachronic histories is by no means novel (Baudouin de Courtenay 1895; see Anderson 1985). This idea has never gone away, though it has enjoyed varying degrees of popularity in the century or so that has passed since phonology began to emerge as a discipline. The review presented here does not aspire to historiographic completeness. In particular, I will largely ignore works in the comparative-historical linguistic tradition (where the validity of appealing to the past to explain the present has generally been taken for granted). Rather, the goal is to introduce the reader to those explanatory models of a diachronic-evolutionary bent which have figured most prominently in the literature on phonology and phonetics—the listener-based sound change model of John Ohala and the Evolutionary Phonology model of Juliette Blevins—and to highlight and summarize in broad strokes the ongoing debate about the potential role of diachronic explanation in phonology.

2. Phonetics in phonology: synchrony or diachrony?

As a simple illustration, let us consider a concrete example of a common sound pattern with a plausibly phonetic explanation: the neutralization of lexical stop voicing contrasts like /t/ : /d/ in pre-consonantal position, whereby /d/ becomes devoiced to [t]. This has an obviously plausible phonetic explanation in terms of speech perception (Steriade 1999). To the extent that stops in this position are not released into a following vowel, there are few and relatively weak perceptual cues available in this position on the basis of which a given alveolar stop can be reliably identified correctly as either [t] or [d], as compared to other positions where the /t/ vs. /d/ contrast is not neutralized (e.g., where a stop is released into a following vowel). If we wish to explain why this sound pattern is so widespread in the world’s languages, why voicing neutralization should preferentially target this kind of position, and why no languages appear to selectively neutralize voicing contrasts in the complement set of environments (e.g., in prevocalic position), several options are available to us. First of all, we might contend that the apparent correlation with perceptual salience is an irrelevant curiosity, and that such asymmetric ‘licensing’ of phonological contrasts is defined in abstract-structural terms which are themselves entirely substance-free (for an approach along these lines to sound patterns involving nasality, see Ploch 1999). On this view, the observed asymmetries and typological gaps are a consequence of the inherent properties of an innate Universal Grammar: languages with selective voicing neutralization in prevocalic position are simply impossible systems.

A second alternative is to accept the notion that the phonetic-perceptual factors are what explains the occurrence and prevalence of phonological patterns of this type, and that this phonetic basis of phonological patterns and processes is an intrinsic aspect of the synchronic grammars of languages. Though this position in no
way presupposes an optimality-theoretic conception of grammar (cf. Stampe 1973; Archarangeli & Pulleyblank 1994), we may express it in optimality-theoretic terms as the assumption that the well-formedness constraints which encode synchronic phonological knowledge are themselves motivated (and perhaps even defined) in explicitly phonetic terms. There are two versions of this view. One is that the constraints in question are part of an innate constraint inventory specified by UG, and that the set of constraints contained in UG is ‘grounded’ in phonetics (Kager 1999:11). Another possibility is that constraints are not innate but constructed by the language learner, who draws upon a general and grammar-external body of ‘phonetic knowledge’ (Kingston & Diehl 1994) that informs and restricts the space of constraints that learners will posit (Hayes 1999; Hayes & Steriade 2004; see also Smith 2004). Returning to our voicing neutralization example, the fact that such neutralization often selectively targets pre-consonantal but not pre-vocalic position is a direct consequence of the constraints which make up synchronic grammars. Grammars may contain constraints specifically penalizing contrastive voicing in hard-to-perceive positions (or ones requiring contrast preservation in easy-to-hear position, if voicing as such is penalized), either as a reflection of UG or through ‘induction’ from general phonetic knowledge by language learners. Grammars with constraints which express preferences diametrically opposed to these are impossible, such constraints are not innately supplied in UG, nor do they conform with learners’ implicit knowledge.

The third alternative approach—the focus of this review—is to acknowledge the functional bases of phonological patterning in the phonetic factors of speech production and perception, but to place the locus of phonetic influence squarely in the diachronic dimension of language change, rather than in the synchronic-universal domain of grammars and their ‘design’. Viewed from this perspective, recurrent sound patterns are the product of recurrent diachronic events (sound changes), which have their ultimate causes in the physical conditions under which speaker-listener interactions take place in language use and language transmission. On this view, voicing is not neutralized in preconsonantal (as opposed to prevocalic) position because some constraint to this effect is part of the innate endowment of humans, nor because learners are predisposed to posit only constraints which are phonetically ‘grounded’. Rather, languages will show some tendency to acquire such neutralization patterns for the simple reason that, in positions where distinctive voicing is hard for listeners (including learners) to detect, listeners/learners will be liable not to detect it, erroneously interpreting a preconsonantal voiced obstruent as being voiceless and encoding it as such in their mental representation of the word-form in question. When the pattern caused by such recurring misinterpretations becomes entrenched, the result is a language with systematic voicing neutralization precisely in those positions where it is phonetically motivated. Languages with the opposite, phonetically unmotivated pattern (neutralization in prevocalic position) will be impossible—or at least infrequent to the point of being unattested—to the
extent that this pattern cannot be attributed to plausible diachronic trajectories in terms of sequences of known types of events (sound change, analogical change).

3. The listener-based model of sound change

One of the most influential conceptualizations of how phonetic factors influence sound patterns through diachronic change is due to John Ohala and his listener-based theory of sound change (Ohala 1981, 1989, 1993, and numerous other works). According to this theory, sound change occurs when a listener misperceives or misparses the acoustic signal produced by the speaker, arriving at a representation which differs in some respect from that intended and encoded by the speaker. The phonologization (Hyman 1977) of such misapprehensions on listeners’ part thus provides a channel through which articulatory, aerodynamic and acoustic-perceptual factors come to shape phonological systems. While Ohala is rarely explicit on this point, misperception would be most likely to take place in words with which the listener is relatively unfamiliar, and phonologization through misperception is therefore perhaps most likely to occur at the stage of language acquisition, though to a certain extent lexical acquisition continues throughout the lifespan of the adult language user as well (as do changes in pronunciation, cf. Harrington et al. 2000; see §5 below).

There is little doubt that something along these lines is what accounts for A > B sound changes in cases where A and B are acoustically similar but articulatorily discontinuous. These would include such context-free shifts in place of articulation as /θ/ > /f/ or /kʷ/ > /p/, as well as the intriguing phenomenon of ‘rhinoglottophilia’ whereby aspiration shifts to nasalization or vice versa (Matisoff 1975; Ohala 1975). More importantly, however, Ohala’s theory extends as well to context-sensitive sound changes, and in general to a wide array of changes which have traditionally been attributed either to aimless ‘articulatory drift’ (Paul 1880) or to the complementary, and often conflicting, teleological goals of minimization of (articulatory) effort and maximization of (perceptual) clarity (Grammont 1933). It is generally recognized that in the act of speech perception, the listener is constantly engaged in normalization, correcting for predictable variation in the acoustic signal in order to arrive at the representation intended by the speaker. To mention but one example, Beddor et al. (1986) show how, while nasalization has a lowering effect on the perceived quality of a vowel, listeners are able to factor out this lowering effect when there is a plausible contextual source present in the form of an immediately adjacent nasal consonant. Ohala’s fundamental insight is to attribute a variety of sound changes to such normalization gone wrong, as it were, where a listener either fails to correct for a contextual effect or wrongly attributes some intrinsic property of a segment to mere contextual influence.

Let us consider in some detail an example where the underlying phonetic factors are arguably aerodynamic (Ohala 1983). When a voiceless stop is released into a following vowel or sonorant, the Voice Onset Time (VOT) typically depends on the degree of constriction of the following segment, and so does the amount of frication
noise in the release burst itself. Thus, for example, we will normally find noisier bursts and longer VOT values in sequences like [tɹ] or [ti] than in, say, [te] or [ta]. The aerodynamic explanation for this state of affairs rests on the fact that in order for the vocal folds to vibrate (as in voicing), air must be able to pass through the glottis with sufficient velocity, and this is in turn dependent on there being a large enough drop in air pressure from the subglottal cavity to the oral cavity. During the closure phase of a stop like [t], oral pressure builds up until it equals subglottal pressure, but when the closure is released and air escapes, oral pressure falls abruptly. How fast it falls, and how long before the subglottal-oral pressure differential has reached the critical threshold value at which voicing becomes possible, will depends on the width of the channel through which air has to escape out of the mouth. Consequently, the narrower the constriction into which the stop is released, the longer it will take for enough air to have left the oral cavity so that voicing can commence; furthermore, the narrower the channel the more turbulence will be generated as air flows through it. Hence a [t] will display a longer VOT and a noisier release burst when it is released into a voiced approximant, glide or high vowel than when released into a mid or low vowel. This predictable effect can make (plain, unaspirated) stops sound somewhat like aspirated stops or affricates in this environment. When a listener fails to blame this predictable phonetic effect on the context, and instead perceives it as an intrinsic property of the stop (as either aspiration or affrication), a sound change has occurred (technically, a ‘mini’ sound change; see below for clarification). This, then, is what accounts for common sound patterns whereby stops are obligatorily aspirated or affricated before high vowels and glides (e.g., in Japanese, where /tuw/ → [tsɯ], /ti/ → [tʃi], /tjV/ → [tʃV]).

As mentioned above, there are two ways in which a listener may be mistaken in her decoding of the acoustic signal: by failing to correct for a contextual effect, or by over-correcting, attributing an intrinsic property to contextual influence. This translates into a two-way typology of listener-induced sound changes. In hypocorrective change, a contextual effect is misinterpreted as an intrinsic property of the segment, segment sequence, or word in question. The most obvious instances of this process are assimilatory sound changes, for example in vowel harmony (as well as umlaut, metaphony, etc.), which are the phonologized reflection of earlier V-to-V coarticulation patterns (Ohala 1994; see Przedziewski 2005). Numerous other phenomena also fall under the hypocorrection rubric, such the development of aspiration and/or affrication before high vowels and glides, as described above. Tonogenesis (the emergence of distinctive tones) originates in the well known phonetic effect of contextual pitch perturbations conditioned by the laryngeal properties (e.g., voicing) of an adjacent consonant (Hombert et al. 1979). The emergence of ‘excrecent’ stops in cases like English prin[t]ce (cf. also Thompson < Thom+son) is due to coarticulatory overlap of the oral closure gesture (for the nasal) and the velic closure gesture (for the following [s]), which creates the percept of an intervening oral stop homorganic with the nasal (Ohala 1981). Similarly, articulatory overlap can mask the percept of the middle consonant in a CCC
sequence, such that a phrase like *perfe/kt m/emory* is misperceived as *perfe[k m]emory*, even when the ‘deleted’ /t/ is in fact being articulated by the speaker (Browman & Goldstein 1990)—a likely explanation for consonant deletion in such environments. Compensatory lengthening can likewise be attributed to hypercorrective sound change through the phonologization of pre-existing subphonemic differences in vowel duration (and/or in the duration of V-C transition cues), the phonetic sources of which may be quite varied (Kavitskaya 2002).

The other half of Ohala’s sound change typology is *hypercorrective* change. In situations where the context provides a plausible source for some phonetic property of a segment’s realization, the listener may erroneously ‘undo’ this aspect of the segment. The most obvious instance of hypercorrective change is dissimilation. For example, when a phonemically labialized consonant is adjacent to a rounded vowel, a listener may attribute the labialization on the consonant to coarticulatory influence from the adjacent vowel, and interpret it as a (phonologically) non-labialized consonant. A well-known case is Classical Greek (/lukos/ < */lukʷos/ ‘wolf’, /kuklos/ < */kʷukʷlos/ ‘wheel’); the deletion of postconsonantal /w/ before rounded vowels in English has a similar explanation (*sword* /sɔrd/ < /swoːrd/, *two* /tu/ < /twː/). A notable aspect of this theory of dissimilation and other hypercorrective sound changes is that they are predicted to be structure-preserving (Ohala 1993). In order for a listener to interpret [A] as being [B] + coarticulation, a [B] must already exist in the system (presumably reflecting /B/).

Two points about Ohala’s listener-based model are worth making here. Firstly, it should be emphasized that Ohala’s explanations pertain, strictly speaking, only to the *initiation phase* (‘actuation’) of sound changes. The misperception events on which the model relies, and which have been replicated in the laboratory in a variety of experiments, take place in a particular listener’s perception of a particular utterance during some particular communicative event. The theory has nothing to say about how such a ‘mini sound change’ takes hold and spreads throughout a community. Nor does it account for how the effects of this ‘mini sound change’ come to be manifested across the entire lexicon rather than just in the one word in question. With respect to the latter question, proponents of Ohala’s model typically appeal to some version of *lexical diffusion*, whereby sound changes are assumed to spread through the lexicon one word at a time (Wang 1969; Phillips 1984; Labov 1994). On this view, the exceptionlessness so characteristic of sound changes (as per the well known ‘Neogrammarian principle’) merely represents an endpoint of a gradual process of lexical diffusion (Krishnamurti 1998).

Secondly, the explanation of a sound change A > B in terms of perceptual error, where [A] (in some context) is misperceived as [B], does not imply that sound changes of the reverse type (B > A) ought to be equally possible. In some cases, such symmetry is expected; /kʷu/ > /ku/ (through hypercorrection) and /ku/ > /kʷu/ (through hypocorrection) are both possible, and both types of sound change are indeed attested. However, as has been amply revealed in perception studies, perceptual confusion is very often asymmetric. For example, Guion (1998) reports
that when stimuli were partly masked by white noise, misidentifications of /ki/ as /tʃi/ occurred three to four times as often as the converse confusions (/tʃi/ heard as /ki/), which were quite rare (cf. Winitz et al. 1972; Plauché et al. 1997). Consequently, the bizarre and unattested change /tʃi/ → /ki/ is predicted to be impossible (or at least extremely unlikely). The unsupported assumption that misperception necessarily be symmetric is occasionally encountered in works critical of phonologization accounts in terms of listener-based sound change (e.g., Steriade 2001; Kiparsky 2004).

The greatest success in the research program initiated by John Ohala and continued by many of his students and associates has been the identification of articulatory, acoustic and/or aerodynamic factors underlying a great number of common and less common synchronic sound patterns. In particular, these results have led to an increased awareness on phonologists’ part of the fundamental role which speech perception has to play in explaining such patterns and processes (see, e.g., many of the contributions in Hume & Johnson 2001 and Hayes et al. 2004).

4. What is the role of the speaker in sound change?

In sharp contrast to the traditional view in historical linguistics, Ohala’s model assigns the speaker at best a tangential role in initiating sound changes and in influencing the direction of such changes. In the strongest version of the listener-based theory, the speaker’s only role is to contribute to the ‘pool of synchronic variation’ (Ohala 1989) from which the listener draws her conclusions about what mental representation underlies the speech signal. There are reasons to believe that this view may be overly simplistic, and that speech production factors play a somewhat more direct role in shaping sound change—and hence, by transitivity, the synchronic sound patterns resulting from such change.

For example, the typology of recurrent sound changes is not always consistent with perception studies in the way that one would expect if all phonetically based sound patterns are ultimately due to misperception. Consider, for example, the extremely common process of nasal place assimilation, whereby nasals take on the place of articulation of an immediately following consonant (/m+d/ → [nd], /n+k/ → [ŋk]). This is consistent with the relative weakness of acoustic cues for place of articulation in this environment, particularly in nasals (Ohala 1990). Indeed, Hura et al. (1992) find that in heterorganic VC₁#C₂V sequences, where C₁ ranged between nasal, fricative or stop, C₁ place is misidentified far more often than C₂ place, and most often when C₁ is a nasal. Misperception is thus a plausible explanation for the regressive directionality of assimilation, and for the particular susceptibility of nasals to such assimilation. However, the assimilatory character of the process is itself not supported by the experimental findings: when subjects mishear the place of articulation of C₁, they typically perceive it as consistently alveolar, regardless of C₂ place. Hura et al.’s (1992) alternative interpretation of assimilation in such clusters is that it constitutes ‘perceptually tolerable articulatory simplification’ on the part of the speaker, not misperception of a heterorganic
cluster as homorganic on the part of the listener. This interpretation is set against Lindblom’s (1990) ‘H&H theory’ of variation in speech production, in which the speaker is tacitly aware of factors which may affect the listener’s ability to correctly perceive the signal, and actively tunes her own production to compensate for those factors (Lindblom et al. 1995; see also Steriade 2001). On this view, then, the actual change from unassimilated to assimilated cluster is initiated by the speaker, not the listener (though see §5 below for a modified version of this view).

This highlights another important aspect of Ohala’s listener-based model of sound change, namely its fundamentally non-teleological character (see Blevins & Garrett 2004 for recent discussion). In Ohala’s model (unlike that of Lindblom), sound change does not serve any ‘purpose’ or ‘goal’, such as to make a word simpler or less effortful for the speaker to produce or easier for listeners to perceive reliably. Consequently, neither do the resulting synchronic sound patterns reflect ‘optimization’ on such functional parameters in any meaningful sense. Though teleological explanations have traditionally been eyed with great suspicion in historical linguistics (pace Grammont 1933, Martinet 1955; see, e.g., McMahon 1994, Lass 1997), the output-oriented and constraint-prioritizing character of Optimality Theory renders phonological derivations—and hence the application of phonological processes—as inherently goal-oriented in character. Consequently, the vehemently anti-teleological stance inherent in Ohala’s phonologization approach to the phonetic explanation of sound patterns has created a deep conceptual divide between works advocating this approach and ones closer to the mainstream of current generative phonological theory. A good example is the treatment of various positional neutralization phenomena in Barnes (2006) vis-à-vis the Optimality Theory analyses developed in Crosswhite (2001) and Smith (2005).

Though much work remains to be done in this area, various experimental studies appear to support the notion of listener accommodation in speech production. The pronunciation of a word is affected not only by factors like frequency and relative ‘information content’ (contextual predictability; Lieberman 1963), but also lexical neighborhood density. Words from high-density neighborhoods in the lexicon, which are ‘hard’ for listeners from the point of view of word recognition and lexical retrieval, appear to show hyperarticulation effects (Munson & Solomon 2004; Wright 2004) as well as greater amounts of coarticulation (Scarborough 2004). Albeit suggestive, such findings are not always an unambiguous indication that speakers actively strive to make the listener’s task easier. Pierrehumbert (2002) offers an alternative account of Wright’s (2004) findings of vowel hyperarticulation in high-density words, which is couched in an exemplar-based model (on which see §6). Her suggestion is that it is instead listeners who are being selective in preferentially encoding hyperarticulated tokens of such words in episodic memory—which in turn will influence their own future productions of the words in question.

While much is yet unclear about the relative contributions of the speaker and the listener to sound change, one thing has become abundantly clear: the ‘pool of synchronic variation’ in speech production is far from random, but rather intricately
structured and influenced in principled ways by a variety of factors. Any explicit theory of sound change, and any theory which seeks to explain the properties of synchronic phonological systems with reference to the diachronic domain of language change, needs to take into account this intricate interplay of perception and production in speaker-listener interactions.

5. Sound change in Evolutionary Phonology

As part of her Evolutionary Phonology model, Blevins (2004; see also Blevins 2005, 2006a, b; Blevins & Garrett 1998, 2004) outlines a model of sound change which is essentially an elaboration of Ohala’s earlier model, and which in part addresses some of the issues raised above. In this ‘amplified’ model of listener-based sound change, referred to as the CCC model, sound changes come in three basic varieties, labelled CHANGE, CHANCE and CHOICE. While it is acknowledged, Ohala’s hypo- vs. hypercorrection dichotomy plays no explicit role in Blevins’ model.

In both of the first two types of sound change, CHANGE and CHANCE, the listener (re)constructs a phonological representation which differs from that intended by the listener, much as in Ohala’s hypo- and hypercorrective misperception scenarios. However, it is only in CHANGE that some aspect of the acoustic signal is actually being misheard. For example, the listener might hear [v] when the speaker in fact produced [ð] (or, similarly, [ki] might be misheard as [tʃ], [kʷu] as [pu], or breathiness as nasalization). Consequently, CHANGE automatically and immediately leads to in a change in pronunciation of the word in question by the listener (as compared to that of the speaker).

The sound changes Blevins labels as CHANCE, by contrast, are ones in which the acoustic signal, and the articulatory events it reflects, is itself genuinely ambiguous in some way, and where the listener simply happens to parse the signal in a way that diverges from the speaker’s intended representation. For example, the speaker may produce /...kʷu.../ as [...kʷu...], with phonemic labialization on the velar stop, whereas the listener, correctly hearing this [...kʷu...], incorrectly parses it as /...ku.../, erroneously attributing the labialization of the [kʷ] to the following rounded vowel (exactly as in the discussion of hypercorrective change in §3 above). It is important to note that this misinterpretation need not noticeably affect the listener’s own pronunciation of the word, as she is likely to continue to render /...ku.../ as [...kʷu...], at least under most circumstances; CHANCE is thus largely covert. In the vast majority of cases, CHANCE involves features or segments which have temporally elongated cues or are otherwise difficult to localize within the segmental string. It is therefore assumed to be the main factor in the emergence of various feature-spreading processes, as well as many instances of metathesis and dissimilation. For example, in Tsilhqot’in, which contrasts pharyngealized and non-pharyngealized alveolar sibilants (e.g., /sˁ/ vs. /s/), and in which uvulars and pharyngealized sibilants alike cause retraction and/or lowering of nearby vowels, all verb stems of the shape -S³VQ have been reanalyzed as -SVQ (see Hansson 2007).
In other words, a pronunciation like [-tsˁɑχ] ‘sinew’, as a realization of /-tsˁæχ/, is instead interpreted as representing /-tsˁæχ/.

The third type of sound change, **choice**, is where Blevins’ model departs from Ohala’s, and where the speaker is implicated to a certain extent. Blevins adopts certain aspects of the H&H theory of Lindblom (1990), viewing variation in production as falling largely on a continuum from relatively hypo-articulated variants to relatively hyper-articulated ones. A speaker thus produces, and a listener is exposed to, a range of multiple variant pronunciations of a single word form. For example, the realizations of /...ut'.../ might range from the relatively ‘hyper’ [...]ut'...], via intermediate [...]yʔt...], [...u't...], [...y't...], to the relatively ‘hypo’ [...]yʔ...], with shorter vowel duration and hence greater centralization/fronting and coarticulatory overlap, as well as near-complete reduction of the oral gestures of the /t'/. If the relative frequency distribution of these variants changes, listeners may come to choose a different phonological representation to represent this continuum. (In a reference to exemplar-based models (see §6), Blevins uses the terms ‘best exemplar’ or ‘prototype’ here, though most such models in fact do not invoke abstractions of that kind.) For example, if variants with considerable vowel fronting become particularly frequent in the ambient production data ([...yʔt...], [...yʔ...]), the listener might choose /...yt'.../ as the lexical representation, or /...yʔ.../ if variants with (largely) debuccalized /t'/ are particularly common.

Blevins (2004) does not cite Hura et al. (1992) and the problems posed by their finding that typical patterns of misperception in heterorganic consonant clusters fail to match the most typical sound change affecting such clusters, namely assimilation (see §3 above). Following Ohala (1990), Blevins clearly considers the primary mechanism for regressive place assimilation in C₁C₂ clusters to be **change**, i.e. outright misperception, a view which would appear difficult to reconcile with Hura et al.’s findings. She does however note (pp. 116, 118–119) that another contributing factor may be ‘coarticulation’, by which she means the (near-)complete masking of the oral gesture of C₁ through gestural overlap (Browman & Goldstein 1990), such that /...np.../ might be pronounced [...]mnp...] with little or no audible trace of the alveolar closing gesture in the acoustic signal. Blevins seems to have this in mind when she states that ‘**choice** may also be involved’ in some cases of assimilation (Blevins 2004:44). Invoking **choice** here would by definition have to mean that such gestural-overlap tokens, having become gradually more frequent in the ambient distribution of production variants to which listeners are exposed, become the basis for positing a phonological representation different from the original one (/...mp.../ rather than /...np.../). Nevertheless, it would seem that **choice** alone is not sufficient even here, and that in order to get from the dual-gesture [...]mnp...] to the consistently single-gesture [...]mp...] = /...mp.../, **change** needs to be invoked as well (rather than **change**, since the acoustic signal is presumably ambiguous as to the [...]mnp...] vs. [...]mp...] articulatory distinction). In other words, the two alternative mechanisms are either **choice** + **change** or else pure **change** (the latter contradicted by Hura et al. 1992). A third alternative
possibility is to embrace more fully Lindblom’s notion of the ‘hypo-hyper’
continuum of production variants as being generated by deliberate ‘perceptually
tolerable articulatory simplification’ on speakers’ part (see §4)—or at least the idea
that fully-assimilated variants ([...mp...], not just ‘coarticulated’ [...]mnp...]) are
innovated as ‘hypo’ production variants by the speaker—in which case choice alone
would be entirely sufficient. If nothing else, this example demonstrates how difficult
it can be to disentangle the various possible mechanisms of change, and the degree
to which the listener and the speaker are each implicated in those mechanisms,
when individual cases are considered in detail.

Finally, one consequence of including choice in the model is that this explicitly
allows for changes in the phonetic realization of words—e.g., in the relative
frequency of pronunciation variants—over the lifespan of the speaker (for an
amusing example, see Harrington et al. 2000). In other words, it is not assumed that
all sound change necessarily constitutes imperfect learning at the acquisition stage.
This in turn opens the door to word-frequency effects, making it possible to
accommodate the observation that most sound changes affect high-frequency words
before low-frequency ones (at least those sound changes which involve lenition,
reduction and/or increased gestural overlap; Bybee 2001; cf. Phillips 1984). In this
respect, the Evolutionary Phonology framework is closely aligned with exemplar-
based models of speech production and perception, to which we now turn.

6. Exemplar-based models and simulations of sound pattern evolution

Traditional generative models of the phonetics-phonology interface (e.g. Keating
1990; Coleman 1998) take the view that phonological representations (underlying
and surface representations alike) are composed of discrete symbolic elements—
possibly redundancy-free as per some version of Underspecification Theory—which
are mapped or ‘transduced’ onto the continuous/analog/gradient domains of
articulation and acoustics. In recent years, this traditional view has increasingly
been challenged by probabilistic exemplar-based models of speech perception and
production (Johnson 1997; Pierrehumbert 2001, 2002). In such models, lexical
entries are represented by clouds of exemplars in episodic memory; these are
essentially memory traces representing tokens previously encountered in perception
(some decay function is assumed, such that older tokens fade over time). In
perception, exemplars are activated probabilistically in proportion to their similarity
to the token under consideration, which is then categorized accordingly. In
production, a subset of exemplars is selected and a production motor plan is arrived
at by averaging over these exemplars (in proportion to their activation levels).

The production-perception feedback loop inherent in such models entails that
exemplar clouds—of entire word forms, as well as of various categories cross-
classifying such whole-word or whole-utterance representations—can and will be
subject to gradual change over time. Language is thus viewed as a fundamentally
dynamic system, and the emergence and evolution of sound patterns at various
levels of granularity can be analyzed in terms of formal and mathematically explicit
models of these complex dynamics in acquisition, in the competence of the adult speaker, and in speaker-listener interactions. Using agent-based computational simulations, Wedel (2004, 2006) demonstrates how the merger, maintenance and/or transformation of phonological contrasts over time can be derived in exemplar-based terms (see also Pierrehumbert 2001, 2002; a less formal instantiation of the same general approach is outlined by Silverman 2006). Other works which exploit the self-organizing capabilities of complex adaptive systems through agent-based simulations of diachronic change include de Boer (2001) on vowel inventories and Harrison et al. (2002) on vowel harmony; see also Boersma & Hamann (2007) on contrast dispersion. In another recent study, Wedel (2007) shows how positive feedback ('analogical error', in the form of a perception bias favoring previously encountered exemplars similar to the current token) can give rise to regularity across the lexicon. Given such reinforcing feedback, the categorical and regular patterns so characteristic of phonological systems emerge as stable states.

In exemplar-based models of this kind, known functional factors—exigencies of production, perception and processing—can act as constant biases exerting pressure on the system (Pierrehumbert 2001; Wedel 2007). The distribution of exemplars may shift under the influence of such factors, which thereby provides the seeds for potential (larger-scale) sound changes of familiar kinds. Moreover, language-specific frequency distributions may also provide a source of biases of a similar kind, with interesting consequences. For example, Wedel’s (2007) simulations closely replicate Gordon’s (2002) typological observation that whether a language counts CVC syllables as heavy or not is strongly (though non-deterministically) correlated with the ratio of sonorants to obstruents in the inventory of syllable codas.

7. Synchronic universals and the adequacy of diachronic explanation

In the ongoing debate about the viability of diachronic-evolutionary approaches, such as that advocated by Blevins (2004), the central question is to what extent such approaches are adequate for correctly predicting observed cross-linguistic typologies, in particular as regards typological gaps and apparent universals. Models like Evolutionary Phonology explicitly allow for ‘unnatural’ histories: diachronic trajectories which involve the telescoping of sequences of independent sound changes, or analogical processes like reanalysis or rule inversion (Blevins 2005; Garrett & Blevins in press). Such models are therefore fairly permissive in the range of synchronic sound patterns they predict to be possible, though patterns which depend on highly specific and fortuitous sequences of events for their emergence are expected to be quite rare, possibly to the point of being (as yet) unattested.

In many ways, the ability to account not only for phonetically ‘natural’ sound patterns but also more arbitrary ones (the ‘crazy rules’ of Bach & Harms 1972; cf. Anderson 1981) is one of the main strengths of Evolutionary Phonology and related models. Typologically aberrant outliers can be accommodated—rather than ignored or explained away by ad hoc manoeuvres—without undermining the soundness of
the strong cross-linguistic generalizations that such systems so blatantly violate. For instance, Barnes (2006) calls attention to the anomalous unstressed-vowel reduction pattern in Seediq, where /e/ is merged with /u/ (as is /o/), rather than with /i/. As it turns out, the vowel which now alternates between (stressed) [e] and (unstressed) [u] goes back to /ə/ at an earlier historical stage. At the time when the unstressed-vowel mergers took place as phonetically driven sound changes, the change in question was thus not the highly unexpected [e] > [u] but rather the less unusual [ə] > [u]. Similarly, Hyman (2001) argues that Tswana shows evidence of a high-ranked constraint against voiced stops in NC clusters (though see Gouskova et al. 2006 for conflicting phonetic evidence). If correct, this contradicts standard assumptions of markedness and typological variation in Optimality Theory (see chapter 2 of Kager 1999 and sources cited therein). Hyman goes on to demonstrate how this synchronic state of affairs is the end product of a sequence of diachronic changes which are themselves unremarkable.

The most serious criticism raised against diachronic-evolutionary models is not that they permit such phonetically unmotivated and typologically aberrant systems, but rather that they do so all too easily and too indiscriminately. The question is whether there are synchronic patterns which should in theory be diachronically accessible, but which are in fact entirely unattested? Kiparsky (2004, 2006) takes just this position, arguing that universal design principles of synchronic grammar—Universal Grammar (UG) in the standard generative sense—constrain the possible outcomes of diachronic change. De Lacy (2006) refers to this constraining influence of UG on cross-linguistic typologies as 'straitjacket effects'. Both Kiparsky and de Lacy acknowledge that the functional pressures of language use do to a certain extent determine typological generalizations and frequencies, and acknowledge the validity of seeking such diachronic-functional explanations. However, they maintain that certain true universals do exist which cannot be adequately accounted for in this manner, and which must therefore owe their existence to properties of the phonological component of UG. In other words, diachronic explanations alone are not sufficient.

A topic which has figured prominently in this discussion is the typology of final voicing neutralization, and whether such neutralization exclusively favors voiceless over voiced obstruents (as predicted by formal theories of markedness; cf. Lombardi 2001). Lezgian has been claimed to show an active process of syllable-final voicing (Yu 2004), and Blevins (2006a) adduces several additional cases, arguing that the cross-linguistic prevalence of devoicing over voicing in final neutralization is merely a (strong) statistical tendency dictated by recurrent patterns of sound change. Kiparsky (2004, 2006) disputes Yu’s and Blevins’ interpretation of the phonetic and phonological facts of each of these cases, maintaining that the typological gap is in fact absolute and hence a true universal. Moreover, as Kiparsky (2006) points out, it is easy enough to imagine simple sequences of diachronic changes which ought to be able to give rise to synchronic patterns of final voicing rather than devoicing. For example, consider a system with a geminate vs. singleton stop contrast (/matt-a/ :
/mat-a/) which is neutralized by final degemination (\{/mat/, /mat/\} \rightarrow [mat]).
Given a lenition sound change whereby [t] \rightarrow [d] and [tt] \rightarrow [t] in all positions,
and concomitant restructuring of the underlying contrast (such that the old /tt/ : /t/
opposition is now /t/ : /d/), the result will be a system where /mat/ \rightarrow [mad] but
/mat-a/ \rightarrow [mata]; in other words, final neutralization in favor of voiced stops.

Even if one accepts Kiparsky’s argument, and his alternative interpretations of
alleged instances of final voicing, it is not quite clear how one should envisage the
intervention of UG in blocking the emergence of cases like the hypothetical one just
described. One not so plausible interpretation is that UG has a ‘prophylactic’ effect,
quite literally blocking such a (phonetically driven) lenition sound change from ever
taking root in a language with these properties. In an analogous language lacking
final degemination, by contrast, the exact same sound change would presumably
have been allowed to progress unhindered by UG. A more plausible alternative
interpretation is that the lenition sound change would still be able to take place,
producing a sound pattern which would give the appearance of final voicing, but
where the neutralization pattern would have no status as such in the synchronic
phonological grammars constructed by learners. The distributional gap (absence of
/t/ in final position) would thus be treated by learners as an ‘accidental’ gap waiting
to be filled by loanwords and other new formations. Observed [t] \sim [d] alternations
would be acquired in terms of listed allomorphs (e.g., /mat/ \sim /mad/), and should
thus be prone to analogical levelling and decreased productivity over time.

For example, in the Northern dialect of Icelandic, aspirated and unaspirated
voiceless stops contrast word-initially (/tʰau/ ‘toe’, /tau/ ‘coma’), but intervocally
only the aspirated stops occur (/kautʰa/ ‘riddle’; Hansson 2003). The historical
background is well understood: in postvocalic positions, the Modern Icelandic reflex
of the Proto-Indo-European contrast from which the initial /tʰ/ : /t/ distinction
derives is instead /tʰ/ : /ð/. Perhaps unsurprisingly, a steady stream of borrowings
over the past few centuries have gradually been filling the gap (/ratar/ ‘radar’,
/lekou/ ‘lego blocks’). This might be taken as evidence that the ‘neutralization’ in
favor of aspirated stops was never an aspect of the Northern Icelandic phonological
system in the first place, even at the earlier time when the distribution pattern
facing the language learner would have been exceptionless. Variant pronunciations
in some items show that aspiration has occasionally been imposed (e.g. /tʰupa/ \sim
/tʰupʰa/ ‘tube’), which suggests that an account along these lines may be overly
simplistic. In any case, the main problem with this view of UG-diachrony interaction
is that it is hard to falsify in principle; apparent counterexamples can all too easily
be explained away as not belonging to the ‘real’ phonology of the language.

From the point of view of a generative phonologist, true universals, if they exist,
are restrictions on possible grammars—knowledge systems internalized by language
learners. The assumption that a generative phonological module of grammar exists,
particularly if this module is assumed to manipulate symbolic elements which are
transduced into the phonetic domain by some ‘phonetic implementation’ function,
brings with it certain complications. Ambiguities about whether certain descriptive
facts constitute an aspect of the knowledge system (in the generative sense)—or, if they do, ambiguities about how these facts are encoded and represented in the grammar—will by necessity make it much harder to settle disputes about whether marginal cases of typological gaps are genuine gaps or not. For example, a pair of segments in different languages which are phonetically (more or less) the ‘same’ kind of segment may be quite distinct phonological objects in terms of their featural composition, with consequences for how such segments are predicted to pattern. A phonetically ‘voiced’ stop is thus not necessarily a phonologically [voiced] segment (Jessen & Ringen 2002). Blevins (2006b) criticizes Kiparsky (2006) for explaining away some of her examples of final voicing on the grounds that they involve non-phonological voicing in this sense. To the extent that our model of the phonology-phonetics mapping is not wholly deterministic (and in any case rather under-developed)—such that we cannot at present unequivocally ‘discover’ a language’s featural representations from the phonetic signal—such criticism is not without justification. However, as long as the debate about intrinsic universals and their role in explaining the typology of sound patterns is taking place against the backdrop of a symbolic-generative grammar model, ambiguities of this kind must be taken seriously, and attempts at resolving them in a principled way must be sought.

The question whether diachronic-functional explanations are wholly sufficient in accounting for phonological typology, or whether they must be supplemented with (and/or replaced by) synchronic universals attributed to some version of UG, is far from settled. A somewhat independent question is the relative intrinsic merit of the diachronic and synchronic modes of explanation for linguistic phenomena. Blevins states, as the central premise of Evolutionary Phonology, that ‘[p]rincipled diachronic explanations for sound patterns have priority over competing synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted’ (Blevins 2004:23, emphasis added; in a later version Blevins 2006a:124f downplays the synchrony by substituting ‘phonological’ vs. ‘extra-phonological’ for ‘synchronic’ vs. ‘diachronic’). The idea is of course not new (see §1), but the territorial dispute between diachronic vs. synchronic/UG explanations has intensified sharply in the context of Optimality Theory, in which functionally ‘grounded’ constraints are standardly proposed and (usually) attributed to an innate UG, as described in §2.

De Lacy (2006) suggests that there is no reason why the availability of a plausible diachronic-evolutionary explanation for some typological generalization should preclude the existence of a UG-internal explanation as well. While this is true at some level, the ‘priority’ Blevins assigns to diachronic explanation is partly based on an Occam’s Razor argument. Diachronic explanations of the sort advocated by Ohala or Blevins are in a very fundamental sense reductionist: the explanandum (some recurrent sound pattern or typological generalization) is accounted for in terms of an explanans based in a concrete and observable domain which is subject to known physical laws and amenable to direct experimental verification (aspects of physiology, aerodynamics, acoustics, perception). The alternative, to posit some
innate constraint or constraint family as part of UG (or to rely on already-posed constraints in some novel ranking arrangement), locates the explanans in a domain which is itself hypothetical and essentially unobservable. With this in mind, it would seem sound methodology to operate under the assumption that nothing should be attributed to UG except when an adequate diachronic-functional explanation cannot be formulated (cf. Hale & Reiss 2000). Of course, controversies such as that concerning the typology of voicing neutralization largely revolve around the question of when a diachronic account counts as fully ‘adequate’.

Another claim made by de Lacy (2006) is that a model like Evolutionary Phonology is about ‘performance’, whereas generative theories are about ‘competence’, and that since their domains are disjoint there is no inherent conflict and very little overlap. In practice, however, things are not so compartmentalized. A central aim of diachronic-evolutionary models is to provide an account of cross-linguistic typology (including typological gaps and near-gaps), and typology-fitting has always been a major aspect of Optimality Theory as it is standardly practiced. By factorial typology, every proposal for a new constraint in UG is implicitly a claim about typological variation, and typological surveys are frequently the point of departure in optimality theoretic analyses of specific phenomena. Be that as it may, it is certainly true that frameworks like Blevins’ Evolutionary Phonology or Ohala’s listener-based model of sound change are not theories of the mental representation of synchronic-phonological knowledge (i.e. speakers’ implicit knowledge of the sound patterns being explained). For this reason, one must take with a grain of salt Blevins’ bold assertion that one consequence of Evolutionary Phonology is that ‘Markedness constraints are excised from synchronic grammars’ (Blevins 2004:23). As a technical term in Optimality Theory, a ‘Markedness constraint’ merely refers to any constraint which evaluates the structural properties of phonological output representations. A parochial and functionally non-grounded constraints like the *ND suggested by Hyman (2001) is no less an instance of a Markedness constraint than its functionally grounded near-opposite *NC̊. Likewise, the notion ‘Markedness constraint’ does not itself automatically presuppose innateness: even models which take constraints to be invented by the learner, rather than provided by an innate UG, rely on this construct (e.g., Hayes & Wilson to appear). What is at stake here is not the ‘excision’ of ranked and violable output constraints as such from synchronic grammars, but rather a rejection of the general strategy of dealing with typological generalizations by positing universal, innate constraints on the sound shape of words.

8. Markedness and universals as learning biases?

Many of the typological generalizations for which diachronic explanations have been successfully proposed are statistical tendencies. The pervasive recurrence of certain sound patterns reduces to recurrent sound changes with clear phonetic bases, whereas the comparatively rare occurrence (sometimes to the point of being unattested) of certain other conceivable patterns is attributed to their relative
inaccessibility via such recurrent sound changes. As rightly noted by some critics of
diachronic approaches, standard generative models like Optimality Theory have
nothing to say about such relative typological frequencies, since any constraint
ranking permitted by UG is just as ‘good’ as any other (de Lacy 2006). Nevertheless,
some recent works have proposed an alternative conception of innate phonological
knowledge, in the form of learning biases, which do provide a means of dealing
with some typological asymmetries which are gradient rather than all-or-nothing.

Moreton (2007, to appear) attempts to quantify the relative robustness of the
‘phonetic precursors’ to different types of sound patterns which have plausible
diachronic origins in listener-based sound change, and compares this robustness
measure to the cross-linguistic frequency with which those same sound patterns are
attested. Contrasting patterns which relate tone to tone (e.g., in adjacent syllables)
against ones which relate tone to laryngeal features (of an adjacent consonant),
Moreton (to appear) finds that even though the two have ‘precursors’ (interactions
and correlations in the acoustic signal) of approximately equal magnitude,
phonological tone-tone interaction is vastly more common cross-linguistically than
tone-voicing interaction. A similar case is made for height-height interaction vis-à-
vis height-voicing interaction in Moreton (2007). On the assumption that the
magnitude of a phonetic effect should determine the relative frequency of the
resulting phonologized sound patterns, cross-dimension interactions (tone-voicing,
height-voicing) thus seem to be ‘underphonologized’ relative to same-dimension
ones (tone-tone, height-height). Moreton (2007) shows how artificial language-
learning experiments, on the other hand, yield results which mirror the typological
distribution closely: cross-dimension patterns (height-voicing) are learned far less
reliably than same-dimension patterns.¹ In this particular case, the factor governing
ease of learning can be construed as the inherent complexity of the interaction
pattern (cf. Pycha et al. 2003; Wilson 2003). Moreton (to appear) accounts for the
‘underphonologization’ of cross-dimension interactions by proposing a Bayesian
learning algorithm in which the learner is biased against adding new constraints
which interact with many other constraints.

In a similar vein, Wilson (2006) reports asymmetries in artificial learning tasks
where the determining factor cannot be construed as formal simplicity but rather
revolves around ‘phonetic naturalness’ in some sense. Subjects who learned a velar
palatalization pattern [ke] → [tʃe] generalized the [k] → [tʃ] change to [ki]
contexts as well, whereas subjects who learned [ki] → [tʃi] did not generalize this
naturalness as a ‘substantive bias’ on phonological learning. Differential treatment
of equally-unattested onset clusters is another area which has recently been claimed
to show evidence for such substantive bias—namely the greater ‘splittability’ of

¹ Interestingly, the more easily learned patterns include not only vowel height agreement but
also consonantal voicing agreement, which is typologically quite rare in comparison and may be
largely dependent on complex diachronic scenarios for its emergence (Hansson 2004).
obstruent-obstruent clusters like /bd/ than obstruent-sonorant clusters like /bn/ in infixation (Zuraw 2007 on Tagalog) and in epenthesis (Berent et al. 2007 on English). However, such findings are often hard to interpret unequivocally. For example, even though neither /bn/ nor /bd/ occur as onsets in English, they fit into larger-scale categories which are attested (stop + sonorant) and unattested (non-/s/ obstruent + obstruent), respectively.

Another place where substantive constraints may be in evidence is in the diachronic development of the lexicon. Frisch et al. (2004) suggest that the elaborate (gradient) dissimilatory cooccurrence restrictions on consonantal roots seen in Semitic languages have evolved as a result of the cognitive pressures affecting the lexicon incrementally: ‘lexical items that avoid repetition will be easier to process, and so will be favored in acquisition, lexical borrowing, coining novel forms, and in active usage’ (Frisch et al. 2004:221). In an intriguing study charting lexical development from Proto-Indo-European to Latin to Modern French, Martin (2006) demonstrates how the cross-linguistically typical frequency relation between /b/ (more frequent) and /d/ (less frequent) has gradually manifested itself over time. Not only do /b/-initial lexical items show a greater ‘survival rate’ than /d/-initial ones, but they also appear to be favored in in the morphological derivation of new words, as well as in borrowing. For example, even though /d/-initial words outnumbered /b/-initial ones both in Latin and in Classical Greek—largely due to the near-absence of /b/-initial words in Proto-Indo-European—Latin borrowed far more /b/-initial than /d/-initial words from Greek.

The alternative view that universal ‘markedness’ or ‘naturalness’ takes the form of substantive biases on the acquisition of sound patterns, and perhaps on language use, is not incompatible with diachronic-functional approaches. For example, if such biases exist, they would simply be one of many potential sources of ‘external error’ in simulation models like that of Wedel (2007). Blevins (2006b) stresses that her Evolutionary Phonology by no means rules out the potential existence of innate knowledge or cognitive processing effects. However, to the extent that substantive biases exist which duplicate recurrent sound changes with plausible sources in misperception (e.g., the greater propensity for [k] > [tʃ] to take place before [i] than before [e]), this does complicate the task of teasing apart the true explanatory factors underlying specific typological generalizations.

**9. Summary**

What role to attribute to diachronic change in explaining the typology of sound patterns, the characteristic phonetic naturalness of such patterns, and the various parochial details of individual sound systems, is a question which is as old as the discipline of phonology itself. It is probably safe to say, however, that the current debate surrounding the role of diachronic explanation in phonology, and the relative importance of this and other types of explanatory factors, is the most active it has ever been in the history of modern phonological theory. This has partly been brought on by the formulation of elaborate and explicit models of sound change.
Secondly, our picture of the full range of typological variation in the sound systems of the world’s languages has become much clearer and more detailed—helped along in part by the inherently typological orientation of Optimality Theory—and this picture has revealed just how strongly the patterns of cross-linguistic variation seem to reflect phonetic constraints on speech production and perception. Finally, a tremendous growth in experimental and computational approaches and methodologies to phonological problems has shifted the arena of this debate out of the armchair and into the laboratories (real and virtual alike). It will be interesting to see to what degree the mainstream of formal-generative phonological theory will be shaped by these developments over the next decade or two.

References

Berent, Iris, Donca Steriade, Tracy Lennertz and Vered Vaknin. 2007. What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104.591–630.


Zsiga, Elizabeth, Maria Gouskova and One Tlale. 2006. On the status of voiced obstruents in Tswana: against *ND. *Proceedings of NELS 36*, ed. by Christopher Davis, Amy Rose Deal, and Youri Zabbal, 721–734 Amherst, MA: GLSA.