Diachronic explanations of sound patterns

Gunnar Ólafur Hansson
University of British Columbia

(final revision, May 27, 2008)

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, in particular Ohala’s listener-based model and Blevins’ Evolutionary Phonology framework, and the relevance of exemplar-based models of speech 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 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 (simple spreading
and delinking operations 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’ in defining and circumscribing the sound patterns
of languages, and the study of typological variation—modelled as permutations of a
rank-ordered list of constraints—has taken on a far more central role in
phonological theory. 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,
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), namely
that grammar-internal explanations are frequently redundant in that they ignore the
availability of an alternative explanation—often independently motivated, and
functionally well understood—situated 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 (e.g., 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 parallel
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 by way of Universal Grammar. In
short, functional-phonetic explanations are to be sought and defined in the
The diachronic domain, not attributed to the inherent ‘design properties’ of the synchronic grammars internalized by speakers/learners. The position that *diachronic explanations* for phonological patterns are not only a powerful tool, but a preferable way of deriving both the characteristic phonetic ‘naturalness’ of phonology and the occasional ‘unnaturalness’ of certain sound patterns, is the main 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 in its own right. 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 scholarly literature on phonology and phonetics—in particular the listener-based sound change model of John Ohala and the more recent Evolutionary Phonology model of Juliette Blevins—and to highlight and summarize in broad strokes the ongoing debate about the potential role that diachronic explanation has to play 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] (/…VdC…/ → […VtC…], neutralizing with /…VtV…/). 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, /…VdV…/ vs. /…VtV…/). If we wish to explain why this type of 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 (UG): languages with selective voicing neutralization in prevocalic position are simply synchronically 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 itself an intrinsic aspect of the synchronic grammars of languages. Though this position in no way presupposes an optimality-theoretic conception of grammar (cf. Stampe 1979; Archangeli & Pulleyblank 1994), we may express it in an optimality-theoretic context as the assumption that the well-formedness constraints which encode synchronic phonological knowledge are themselves motivated (and perhaps even formulated) 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; cf. Archangeli & Pulleyblank 1994). 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 which learners can 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 position, but never pre-vocalic position, is a direct consequence of the constraints which make up synchronic grammars. Grammars may thus contain constraints which specifically penalize contrastive voicing in hard-to-perceive positions (or constraints which require contrast preservation in easy-to-perceive positions, if obstruent voicing as such is what 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 of phonetics.

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 across generations. On this view, voicing is neutralized in preconsonantal (as opposed to prevocalic) position not because some constraint to this effect is part of the innate endowment of humans, nor because learners are predisposed to posit only such constraints as are grounded in phonetics. 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. If and when the pattern caused by such recurring misinterpretations becomes entrenched, the result is a language with systematic voicing neutralization precisely in those positions where such neutralization 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 originates in the kind of event 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 the listener’s 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 during language acquisition in childhood, 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, and 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 (unconditioned) shifts in place of articulation as /θ/ > /f/, /f/ > /x/ or /kʷ/ > /p/, as well as the intriguing phenomenon of ‘rhinoglottophilia’ whereby aspiration or breathy voice shifts to nasalization or vice versa (Matisoff 1975; Ohala 1975). More importantly, however, Ohala’s theory extends as well to context-sensitive (conditioned) 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 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) demonstrate that whereas 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 (for other examples of such coarticulatory compensation, see Beddor & Krakow 1999, Beddor et al. 2002). 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 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 [tɪ] 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 rapidly it falls, and how long before the subglottal-oral pressure differential has reached the critical threshold value at which voicing becomes possible, will depend 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 an 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 /tɯ/ → [tsɯ], /ti/ → [tʃi], /tjV/ → [tʃV]).

As mentioned above, there are two ways in which listeners may be mistaken in their 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 hypocorrection 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 Przedziecki 2005). Numerous other phenomena also fall under the hypocorrection rubric, such as 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 ‘exscrent’ 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/k m/emory can become misperceived as perfe[k m]emory, even when the ‘deleted’ /t/ is in fact being fully articulated by the speaker (Browman & Goldstein 1990)—a likely explanation for the high incidence of consonant deletion in such environments. Compensatory lengthening can likewise be attributed to hypocorrective 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 /soɹd/ < /swoɹd/, two /tu/ < /twoː/). A notable aspect of this theory of dissimilation and other hypercorrective sound changes is that they are predicted to be structure-preserving (Ohala 1993; Kiparsky 1995; though see Blevins & Garrett 1998:519–520 for arguments against this position). In order for it to be possible for a listener to (mis)interpret an instance of [A] as being [B] + coarticulation, a [B] must already exist independently in the system.¹

¹ As the square brackets around [B] suggest, we are here dealing with ‘B’ as an element that occurs in phonological surface representations in the language. Whether such a pre-existing [B] reflects an independent phoneme /B/ in underlying representations is essentially an orthogonal issue; in principle, [B] might just as well be a predictable surface variant of some other phoneme (it might even be conceivable for [A] and [B] to constitute allophones of the same phoneme). In any case, Ohala’s model of sound change does not necessarily presuppose the kind of underlying vs. surface representational distinction assumed in generative models. It should also be kept in mind that ‘A’ and ‘B’ need not represent individual segments. In the English example cited above (/swoɹd/ > /soɹd/, etc.), a bisegmental C+ [w] cluster (= ‘A’) is being misperceived,
Two points about Ohala’s listener-based theory are worth making here. Firstly, it should be emphasized that Ohala's explanations pertain, strictly speaking, only to the initiation (actuation) phase of sound changes. The misperception events on which the theory rests, and which have been successfully 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 specific to say about how such a ‘mini sound change’ (Ohala 1993) 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 theory typically appeal to some version of lexical diffusion, whereby sound changes are assumed to spread gradually 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 or equally frequent. In some cases, such symmetry is expected; /kʰu/ > /ku/ (through hypercorrective misperception) and /ku/ > /kʰu/ (through hypocorrective misperception) 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) found that when listening conditions were artificially degraded so as to increase confusion rates, misidentifications of /ki/ as /tʃi/ occurred three to four times as often as the converse confusions (/tʃi/ heard as /ki/), which were quite rare (see also 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).

\[ \text{hypercorrectively, as monosegmental C (= ‘B’) plus coarticulatory rounding from a neighboring vowel.} \]

\[ \text{2 Such asymmetric confusion is unremarkable in and of itself, and is well attested in other perceptual domains. For example, in visual letter recognition, one is far more likely to misperceive a ‘Q’ as ‘O’ (by failing to notice the distinctive tail that sets the former apart from the latter) than to misperceive ‘O’ as ‘Q’ (by spuriously introducing that same characteristic feature). Plauché et al. (1997) suggest that the /ki/ vs. /tʃi/ asymmetry is due to precisely such an all-or-nothing feature in the acoustic signal, namely the spectral peak in the 3–4 kHz range that is characteristic of velars. Guion (1998) is skeptical of this explanation, conjecturing instead that the asymmetry is likely due to palato-alveolar affricates being a perceptually more robust category than velars, owing to their longer duration and greater amplitude.} \]
The greatest success of 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 that speech perception has to play in explaining phonological patterns and processes (see, for example, 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 such changes take. 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 listeners draw their conclusions about what mental representation underlies the speech signal. There is reason to believe that in its purest form this view is overly simplistic, and that speech production factors play a somewhat more direct role in shaping sound change—and hence, by transitivity, the typology of synchronic sound patterns that result 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 were ultimately due to misperception. Consider, for example, the extremely common process of nasal place assimilation, whereby a nasal takes 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) found that in heterorganic VC₁#C₂V sequences, where C₁ ranged between nasal, fricative or stop, the place of articulation of C₁ was misidentified far more often than that of C₂, and most often when C₁ was 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 misheard the place of articulation of C₁, they typically perceived 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 the backdrop of Lindblom’s (1990) ‘H&H theory’ of variation in speech production, in which speakers are tacitly aware of factors which may affect listeners’ ability to correctly perceive the signal, and actively tune their own production to compensate for those factors (Lindblom et al. 1995; see also Steriade 2001). On this view, then, the actual change from an unassimilated cluster...
to an assimilated one is initiated by the speaker, not the listener (though see §5 below for a modified version of this hypothesis).

This highlights another important aspect of Ohala’s listener-based theory 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 of that term. 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 means that phonological derivation—and hence the application of phonological processes—is construed as an inherently goal-oriented procedure. 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 current mainstream of 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 by 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. For example, the pronunciation of a word is affected not only by factors like frequency and relative information content (contextual (un)predictability; Lieberman 1963), but also by lexical neighborhood density. Words from high-density neighborhoods in the mental lexicon, which are thus potentially ‘hard’ for the listener from the point of view of word recognition and lexical retrieval, appear to show hyperarticulation effects in production (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 a conjectural 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 below). Pierrehumbert’s suggestion is that it is instead listeners who are being selective in preferentially encoding hyperarticulated tokens of such words in their episodic memory; this selective storage in turn ends up biasing their own future productions of the words in question.

In the usage-based model advocated by Bybee (2001, 2006), the speaker is afforded a considerably more immediate and active role in the initiation of sound change than in Ohala’s listener-based model. Bybee supports Mowrey & Pagliuca’s (1995) claim that most sound changes involve articulatory reduction of one kind or another (in fact, Mowrey & Pagliuca hold this to be categorically true of all sound changes). Such reduction may be substantive, in that the magnitude of one or more
articulatory gestures is reduced, or it may be temporal, in that sequences of articulatory gestures are compressed or overlapped in time. (Note that it is possible for a particular sound change to involve a combination of both types of reduction at once.) This corresponds rather closely to the kinds of gestural reduction and overlap effects that have been documented for casual speech (Browman & Goldstein 1990). The motivation for these sorts of articulatory reductions to occur over time is assumed to be the very general tendency for repeated and highly practiced motor behavior to result in reduction and temporal compression: ‘[w]ith repetition, neuromotor routines become more compressed and more reduced’ (Bybee 2001:78). In this conception of the mechanism by which sound change originates, the listener is clearly sidelined in comparison to the speaker. At most, the listener influences (passively and indirectly) the extent to which compression and reduction will still allow successful communication to occur (recall the Lindblomian notion of ‘perceptually tolerable articulatory simplification’, discussed earlier).

Although Mowrey & Pagliuca (1995) go to great lengths to argue that many apparent examples of acoustically/perceptually motivated sound changes can be reinterpreted in terms of temporal or substantive reduction, there are certain types of change which are hardly amenable to this kind of explanation, and for which Ohala’s listener-based account is almost certainly correct. In particular, this is true of changes which involve abrupt articulatory discontinuities (see §3 above), such as when labiodental [f] turns into velar [x] (or vice versa), or when breathiness or aspiration turns into nasalization. The same applies to phenomena like tonogenesis, in which low-level and largely mechanical pitch perturbations spawn distinctive phonological specifications for tone on a neighboring vowel. It is difficult to see how adopting the notions of temporal and substantive articulatory reduction would help us better understand how and why such changes occur.

Bybee (2001) is skeptical of Ohala’s general account of sound change as misperception, arguing that ‘it is unlikely that hearers who have already acquired the phonetics of their dialect would misperceive already acquired words to the extent that that might cause a sound change’ (p. 75). She does however concede that certain sound changes might originate in children’s misperception or misanalysis of the acoustic signal at the acquisition stage. Most interestingly, though, Bybee argues (following Phillips 1984, 2001) that articulatorily and perceptually motivated sound changes are expected to interact very differently with word frequency, and should therefore leave distinct lexical diffusion ‘footprints’. Given that the compression and reduction inherent in (speaker-based) articulatorily motivated changes is due to the articulatory automation of often-repeated elements, it would follow that such changes should affect high-frequency forms first, and only later spread to forms with lower token frequency. By contrast, perceptually motivated (listener-based) sound changes, which are essentially a form of imperfect learning, ought to affect low-frequency forms first, since the listener/learner has much greater exposure to more frequent forms and is therefore likely to perceive and acquire those correctly. It remains to be seen whether this prediction stands up
to scrutiny, and what its implications might be for analyses of synchronic systems and phonological typologies. In recent years, probabilistic aspects of sound patterns (e.g., variation and optionality, gradient well-formedness judgements, probabilistic phonotactic restrictions, word-frequency effects) have come to figure more and more prominently within mainstream phonological theory (see Coetzee & Pater 2008 for a recent overview).

It should be noted that Bybee (2001:75) also acknowledges that the outcome of prior (speaker-based) articulatory compression and reduction changes may in turn be subject to (listener-based) perceptual reanalysis later on. For example, a word-final vowel + nasal sequence might undergo considerable gestural overlap (resulting in nasalization of a substantial portion of the vowel) and articulatory reduction of the nasal (resulting in shorter duration, and perhaps less-than-complete oral closure). As a result of these articulatory changes, the misperception of the vowel + nasal sequence as a mere nasalized vowel becomes much more likely than otherwise. In practice, the difference between this account and Ohala’s purely listener-based treatment is largely a matter of emphasis. For Ohala, the gestural reduction and overlap is merely present in the ‘pool of synchronic variation’ (Ohala 1989) on which listener-based sound changes operate (i.e. these articulatory properties happen to be found in some subset of the production tokens that make up that ‘pool’); it is only the perceptual reanalysis event that is deemed to deserve the label ‘sound change’. For Bybee, the focus is instead on the processes by which such articulatorily reduced production variants come into existence in the first place; that process is viewed as the more significant ‘sound change’ event, rather than any perceptual reorganization that may ensue. This difference of perspective also relates to what appear to be somewhat different assumptions about the nature of phonological representations and lexical storage. To the extent that this can be inferred from Ohala’s writings, he is for the most part presupposing something akin to a (quasi-)phonemic representation of word-forms (though probably with canonical, contextually predictable allophonic variants spelled out). Bybee’s usage-based model, on the other hand, rests on a rich-storage conception of the mental lexicon, in which ‘each word is represented with a range of variation that corresponds to the actual tokens of use experienced by the user’ (2001:42), such that ‘phonological representations [of lexical items] are based on categorized tokens of use’ (2001:62). We shall return to this issue in the following section in the context of Blevins’ Evolutionary Phonology model, as well as in the discussion of exemplar-based models in §6 below.

To sum up, while much is yet unclear about the relative contributions of the speaker and the listener to diachronic 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 and cross-linguistic typology of synchronic phonological systems with reference to the diachronic domain of language change, needs to take
into account this complex interplay of production and perception in speaker-listener interactions, both in fully competent adults and in children acquiring language.

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 in the previous section. In this ‘amplified’ model of listener-based sound change, referred to as the ‘CCC model’, sound changes come in three basic varieties, dubbed CHANGE, CHANCE and CHOICE. While Blevins does acknowledge Ohala’s hypo- vs. hypercorrection dichotomy, that distinction plays no explicit role as such in her model.

In both of the first two categories of sound change, CHANGE and CHANCE, the listener (re)constructs a phonological representation which differs from that intended by the speaker, 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 outright. For example, the listener might hear [v] when the speaker in fact produced [ð] (or, similarly, [kʲi] might be misheard as [tʃi], [kʷu] as [pu], or breathy-voiced [a] as nasalized [ã]). Consequently, CHANGE automatically and immediately leads to 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 sequence of 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 [...k“u...], with phonemic labialization on the velar stop, whereas the listener, correctly hearing this [...k“u...], incorrectly parses it as /...ku.../, erroneously attributing the labialization on the [k“] to coarticulation with 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 the listener is likely to continue to render /...ku.../ as [...k“u...], 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 acoustic cues or which are otherwise difficult to localize within the segmental string. It is therefore assumed to be the main mechanism responsible for the emergence of various feature-spreading processes, as well as many instances of metathesis and dissimilation.

For example, in Tsilhqot’in, vowels undergo retraction and/or lowering next to both uvulars and pharyngealized sibilants, and such effects on vowel quality provide the only reliable perceptual cues to the pharyngealization contrast in sibilants (see Hansson 2007 and sources cited therein). In an intended sequence like
/…SˁVQ…/, the retracted and lowered quality of the intervening vowel becomes ambiguous as to its source. The listener needs to decide whether to infer (correctly) that the vowel quality is partly due to the preceding sibilant—and hence that the sibilant in question must be a phonologically pharyngealized one—or instead to conclude (incorrectly) that the lowering/retraction is entirely due to the following uvular, and that the sibilant must thus be of the non-pharyngealized kind. In other words, should a pronunciation like [-ts'ɑχ] ‘sinew’ be interpreted as /-ts'æχ/ or as /-ts'æχ'? As it turns out, all morphemes which were originally of the shape -SˁVQ have been historically reanalyzed as -SVQ in Tsilhqot’in (Krauss 1975; see Hansson 2007), in what amounts to a CHANCE-based dissimilatory sound change.

The third type of sound change, CHOICE, is where Blevins’ model departs most clearly from Ohala’s, and where the speaker becomes 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 individual word forms. For example, the realizations of /…ut'/ might range from the relatively ‘hyper’ [...ut'...], via intermediate [...yt'...,...yt'..., ...yt'...], 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. 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 even /...yʔ'/ if variants with (largely) debuccalized renderings of /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, roughly, as [...n̥mp...] with little

---

3 In an explicit reference to exemplar-based models (see §6), Blevins uses the terms ‘best exemplar’ or ‘prototype’ in this context, even though such models typically do not invoke such abstractions at all. On this point, Blevins’ proposal most closely echoes Bybee (2001), who advocates a kind of hybrid between an exemplar model and a prototype model—even Bybee does not appear to presuppose that each word form is somehow characterized by a single (‘prototypical’) phonological representation in the lexicon.
or no audible trace of the alveolar closing gesture in the acoustic signal. Note that the transcription [еньк] is here intended to convey a doubly-articulated nasal, in which the alveolar and bilabial closures are essentially cotemporaneous; in other words, [...نك... ] represents a /n/ + /p/ cluster in which the labial closing gesture for the /p/ starts already at the onset of the nasal, and thus overlaps the alveolar closure of the /n/. Blevins (2004:44) seems to have this in mind when she states that ‘choice may also be involved’ in some cases of assimilation.

To be precise, 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 [...]nk... ] to the consistently single-gesture [...]nk... = /...mp.../, chance needs to be invoked as well (rather than change, since the acoustic signal is presumably ambiguous as to the [...]nk... ] vs. [...]nk... ] articulatory distinction). In other words, the two alternative mechanisms are either choice + chance or else pure change (the latter being contradicted by the findings of Hura et al. 1992). A third alternative possibility would be 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 above). It is at least conceivable that fully-assimilated realizations (i.e. [...nk...], not just ‘coarticulated’ [...]nk...)] 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 on sound change such as those addressed by Bybee (2001), as discussed in §4 above. In this respect, Blevins’ 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; see also Bybee 2001). In such models, lexical entries are represented by clouds of exemplars in episodic memory. An exemplar is essentially a memory trace representing an individual token previously encountered in perception (exemplar models are also known as multiple-trace models; Hintzman 1986); some decay function is assumed, such that older tokens fade over time. In perception, stored exemplars are activated probabilistically in proportion to their similarity to the token under consideration, which is then categorized in accordance with these. In production, a subset of exemplars is selected and a production motor plan is arrived at by averaging over these selected exemplars (in proportion to their activation levels).

The production-perception feedback loop inherent in exemplar-based models entails that exemplar clouds—of entire word forms, as well as of the multitude of categories that cross-classify 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 and usage-based system (Bybee 2001, 2006; Silverman 2006a), 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. For example, Wedel (2004, 2006) uses agent-based computational simulations to demonstrate how the merger, maintenance and/or transformation of phonological contrasts over time can be derived in exemplar-based terms (see also Pierrehumbert 2001, 2002). An instantiation of the same general approach is outlined by Silverman (2006a), who also shows how an evolutionary exemplar model can explain the preservation and even exaggeration of contrasts over time. Silverman (2006a:135–143; 2006b) discusses a particular sound change in Trique, by which lip rounding has spread rightwards across velar consonants ([unga…] > […ugʷa…]). He argues that this development is fundamentally a matter of contrast enhancement: what was once an [uga] : [uda] contrast has been replaced with a more acoustically distinct [ugʷa] : [uda] contrast. (As no [ugʷ] or [ukʷ] sequences existed in the language prior to this development, the increased category separation did not incur any extra ‘cost’ in terms of other contrasts being infringed upon.) Importantly, Silverman rejects any kind of teleological, speaker-oriented interpretation of the mechanism by which this change occurred (cf. the discussion in §4 above) That is, at no point were any speakers actually striving to make their productions more distinctive or easier to parse. Rather, Silverman suggests a listener-based account ‘by which contrasts might be enhanced passively, evolving over generations of speakers, due to the communicative success of some tokens, and the communicative failure of others’ (Silverman 2006b:141).

The core of Silverman’s proposal is that those tokens of /uga/ which have extensive coarticulatory rounding ([ugʷa]) are more likely to be successfully
perceived and categorized correctly than ones with little or no coarticulation ([uɡa]); a certain percentage of the latter will be misperceived/miscategorized, and will therefore fail to be added to the exemplar cloud representing /uɡa/. This very small but persistent bias will continue to assert itself over cycles of speaker-listener interactions and across generations of learners. Over time, [uɡʷa] exemplars will thus gradually gain ground at the expense of [uɡa] tokens, until they come to dominate the exemplar cloud, at which point they effectively constitute the norm. By contrast, tokens of /uda/ with coarticulatory rounding on the intervening consonant ([uᵈʷa]) fare worse than ones with less rounding ([uda]), as they are less likely to be perceived and categorized correctly (i.e. more prone to being misheard as [uɡa]); therefore, the same mechanism will have the effect of inhibiting excessive coarticulatory rounding in non-velar consonants.4

In response to what she views as problems for simple exemplar-based models of lexical representation, Bybee (2001:138–143) outlines what she calls a ‘modified’ exemplar model. She notes that contextual variants of words are typically not very stable, and that reorganization of variant distributions may occur. For example, in dialects of Spanish where [s] alternates (variably) with [h] and ∅ as realizations of /s/ in pre-consonantal (and to a more limited extent pre-pausal) environments, one may find (e.g. in Cuban Spanish) that in words with final /s/, the frequency of variants with [h] or ∅ has increased dramatically in __#V contexts, under the influence of __#C contexts (in which [h] and ∅ variants dominate), while still remaining quite rare in pre-pausal contexts. The influence exerted by the __#C pattern is clearly related to the fact that the vast majority of Spanish words begin in a consonant; an /s/-final word will therefore find itself in a __#C environment more than twice as often as it will occur in a __#V environment. Bybee notes that if exemplars are stored with their contexts specified, the occurrence of (a particular type of) exemplar in the ‘wrong’ environment is unexpected, and argues (2001:142) that ‘[t]he stabilization of a single variant for a word suggests that representations are exemplar-based, but that all exemplars are not equally accessible’. Her (somewhat sketchy) proposal is that (sub)sets of similar exemplars—regardless of the contexts from which these may be drawn—are mapped onto a single representation. This effectively maps the exemplar cloud to a relatively small set of exemplar ‘types’, which are in principle context-free, and some of which are more frequent overall than others. These ‘more central’ exemplar types ‘are more accessible and may replace the more marginal ones’ over time (ibid.). In the Cuban Spanish example above, [h] variants of words with final /s/ were more central/frequent

4 Though he does not explicitly acknowledge this parallel, Silverman’s (2006a, 2006b) account of Trique labialization is essentially identical to Pierrehumbert’s (2002) suggested explanation for the hyperarticulation of words with many lexical neighbors (see §4 for discussion). In both cases, more ‘exaggerated’ tokens are more likely to be categorized correctly, and will thus gradually—through accumulation over successive cycles of speaker-listener/learner interactions—end up being overrepresented in the exemplar cloud for the category in question.
overwhelmingly than [s] variants were; as a result, the former have ended up infiltrating the
__#V contexts in which [s] variants had been predominant.\(^5\)

Though the question of such ‘variant reorganization’ certainly deserves attention
and explicit analysis, it is not so obvious that the sorts of developments that Bybee
discusses cannot be accommodated in standard exemplar models. The issue would
appear to hinge on the question of exactly how individual exemplars are cross-
classified and categorized along a number of intersecting parameters, and how that
cross-classification is accessed and made use of in production. Exemplar-based
production models like that of Pierrehumbert (2001, 2002) typically assume that in
the production of a given category in a given environment, stored exemplars of that
category from any environment can in principle be activated and contribute to the
calculation of a production target (with exemplars from more similar environments
perhaps being more strongly activated than others). For this reason, it seems that a
mechanism is already in place by which the more numerous __#C exemplars (among
which [h] predominates) can gradually come to influence the incidence of [h]
variants in __#V contexts as well.

In fact, this is exactly how exemplar-based models account for the otherwise
puzzling phenomenon of incomplete neutralization (Port & O’Dell 1985, Warner et al.
2004) and the related issue of near-mergers (Labov et al. 1991, Yu 2007). For
example, in their study of neutral vowels in Hungarian vowel harmony, Benus &
Gafos (2007) demonstrate how the phonetic realization of neutral [i, iː, eː] in those
monosyllabic roots which idiosyncratically take back-vowel suffixes ([iːr] ‘write’, cf.
[iː-r-ɔk] with the dative suffix) is slightly but significantly more back than that of
the same vowels in analogous front-harmonic roots ([h[iː]r] ‘rumor’, dative [h[iː]-
ŋk]). This is true even in isolation, when no suffix follows the roots in question.
The explanation for this subtle and subphonemic ‘contrast’ is that in the production
of an unsuffixed form like [iːr], exemplars from suffixation contexts (e.g. [iː-r-ŋk])
are activated as well. In those latter tokens, the [iː] is subject to coarticulatory
backing due to the nearby back vowel, and therefore has a more retracted
realization. Under the influence of such retracted-[iː] exemplars, the average
production of a back-harmonic root like [iːr] in isolation gets shifted towards a
slightly more back version of [iː] (as compared to that of a front-harmonic root like
[h[iː]r]). The same explanation applies to the incomplete neutralization frequently
observed in word-final devoicing processes in languages like German or Dutch (Port
& O’Dell 1985, Warner et al. 2004). The phonetic realization of the word-final stop
in German Rad ‘wheel’ (traditionally transcribed [ʁaːt]) is influenced by the voiced
realization in exemplars of that same word from suffixed contexts (e.g. genitive Rad-
es [ʁaːdəs]), such that this devoiced ‘[t]’ is in fact not quite phonetically identical to
the genuinely voiceless [t] of Rat ‘counsel’. If exemplar-based models are able to

\(^5\) However, it seems that this account does not provide an explanation for why the spread of [h]
variants should target (phrase-medial) __#V contexts specifically, to a much greater extent than
it did (phrase-final) pre-pausal contexts.
accommodate this sort of cross-context transfer, then they should have no trouble dealing with the kinds of diachronic changes that lead Bybee (2001) to propose what amounts to a hybrid between an exemplar model and a prototype model.

In exemplar-based models of the kind outlined above, known functional factors—exigencies of production, perception and processing—can act as constant biases exerting pressure on the system over time (Pierrehumbert 2001; Wedel 2007). The distribution of exemplars may shift under the influence of such factors, which in turn 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, the results of the computational simulations by Wedel (2007) turn out to replicate closely Gordon’s (2002) typological observation that whether a given language categorizes CVC syllables as phonologically heavy (bimoraic) or light (monomoraic) is strongly, though non-deterministically, correlated with the ratio of sonorants to obstruents in the inventory of syllable codas in the language in question. In other words, the more sonorous a language’s ‘average’ coda is, the more likely that language is to attribute weight to all codas, regardless of their sonority.

The full explanatory potential of exemplar models is most profitably investigated with the help of computational implementations, often in the form of agent-based simulations, which help explore how sound patterns, representations and contrasts are expected to evolve over time in such models, and to what extent these predictions match what is known about attested diachronic changes and synchronic-phonological patterns and typologies. The most notable representative of this strand of research is recent work by Wedel (2004, 2006, 2007). Other works which exploit the self-organizing capabilities of complex adaptive systems through agent-based simulations of diachronic change (though not necessarily explicitly couched in exemplar-based terms) include de Boer (2001) on vowel inventories and Harrison et al. (2002) on vowel harmony; see also Boersma & Hamann (2007) on contrast dispersion, which relates to the work on contrast maintenance and enhancement already cited (Silverman 2006a, b; Wedel 2004, 2006). In a 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 end up emerging as stable states. This line of inquiry, whereby higher-order organizational aspects of synchronic phonologies are themselves viewed as emergent structures arising from the cumulative interaction of other, more basic factors, is likely to yield more interesting results in the future.

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 explanatory frameworks. 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 unstressed /e/ is merged with /u/ (as is /o/), rather than with /i/. As it turns out, the vowel which now alternates between (stressed) front unrounded [e] and (unstressed) back rounded [u] goes back to central /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 nasal+stop clusters, the effect of which is diametrically opposite to the extremely common (and ‘natural’) process of postnasal voicing in the world’s languages. If correct, this contradicts standard assumptions of markedness and typological variation in Optimality Theory (for example, 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 particular sequence of diachronic changes which are themselves quite 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 categorically unattested. Kiparsky (2004, 2006) takes just this position, arguing that a set of universal design principles of synchronic grammar—Universal Grammar (UG) in the standard generative sense—constrains the possible outcomes of diachronic change. De Lacy (2006) refers to this constraining influence of UG on cross-linguistic typologies as ‘straitjacket effects’ (see also de Lacy & Kingston 2006). Both Kiparsky and de Lacy accept that the

---

6 Zsiga et al. (2006) cite conflicting phonetic evidence and call into question Hyman’s (2001) analysis of Tswana in terms of a *ND constraint. However, the existence of a postnasal devoicing process in Tswana is confirmed by Coetzee et al. (2007).
functional pressures of language use do to a certain extent determine typological
generalizations, statistical tendencies and frequencies, and acknowledge the validity
of seeking such diachronic-functional explanations. However, they maintain that
certain *true* universals also 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 capable of giving rise to synchronic patterns of final voicing rather than
devoicing.

For example, consider a hypothetical language which originally had a geminate
vs. singleton contrast in stops, such that /matt-/ and /mat-/ were distinct
morphemes (e.g. realized as [matt-a] vs. [mat-a] when some suffix /-a/ is added).
Let us further suppose that in this language, the length contrast was systematically
neutralized by degemination in word-final position, such that unsuffixed /matt/ →
[mat], surfacing identically to underlying /mat/ → [mat]. Now, if this language
were to then undergo a lenition sound change whereby [t] > [d] and [tt] > [t] in
all positions, and concomitant restructuring of the underlying contrast—such that
the old /tt/ vs. /t/ opposition is now /t/ vs. /d/ instead—the result would be a
system with word-final neutralization of the /t/ : /d/ contrast in favor of voiced
stops. A root like /mad-/ (reflecting earlier /mat-/) would retain voiced [d] in all
positions (suffixed [mad-a], unsuffixed [mad]). On the other hand, a root like
/mat-/- (going back to earlier /matt-/) would display a [t] ~ [d] alternation: the
stop would surface as voiced [d] in final position (unsuffixed [mad]) but as
voiceless [t] elsewhere (suffixed [mat-a]).

Even if one accepts Kiparsky’s argument, and his alternative interpretations of
alleged instances of final voicing, it is not immediately obvious 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 place 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. To the
best of my knowledge, this is not a commonly held version of the view that synchronic typological universals exist independently of diachronic considerations. A less extreme version of the view that synchronic universals constrain diachronic change is that the lenition sound change in question would still be able to take place, but that the resulting sound pattern (which would give the appearance of an active word-final voicing process) would either automatically trigger further changes or simply not be acquired by learners as an aspect of the synchronic phonological grammar they construct. For example, the resulting distributional gap (absence of /t/ in final position) might thus be treated by learners as an ‘accidental’ gap, waiting to be filled by loanwords and other new formations. Observed [t] ~ [d] alternations (as in [mat-] ~ [mad-]) would be lexicalized, dealt with by the learner in terms of listed allomorphs at the underlying level ([/mat-/, /mad-/]), and should thus become highly susceptible to analogical levelling and decreased productivity over time.

Building on work by de Lacy (2002, 2004) on universal markedness relations in sonority-driven stress systems, Kiparsky (2004) argues that in such a stress system, a simple and seemingly innocuous sound change ought in principle—if synchronic properties of phonological systems were merely a product of their diachronic histories—to be able to give rise to a markedness reversal that subverts otherwise exceptionless typological universals. For example, consider a stress system like that of Gujarati (de Lacy 2002), in which stress normally falls on the penultimate syllable ([apwána] ‘to give’, [ekóter] ‘71’), but is attracted away to the antepenultimate or the final syllable if the vowel in that syllable is more sonorous than the one in the penult ([tádʒetar] ‘recently’), where sonority is defined in terms of the (partially conflated) hierarchy [a] > [e, o, i, u] > [ə] (with ‘>’ standing for ‘is more sonorous than’). Kiparsky (2004) points out that a simple context-free sound change [a] > [ə] (such as occurred in Sanskrit, for instance) would be capable of turning such a system on its head. As a result, [ə] might come to attract rather than repel stress, and would effectively end up counting as the most sonorous vowel for the purpose of stress assignment. However, no stress system with such properties exists, and in Kiparsky’s view this is because it would constitute a synchronically impossible system. Importantly, he does not deny the possibility that an [a] > [ə] sound change could occur in a language like Gujarati, but rather claims that this would ‘destroy the phonological regularities of Gujarati’s stress system, with the result that it would have to be reanalyzed with lexically marked stress’ (Kiparsky 2004:11). In other words, learners confronting the resulting surface patterns of stress assignment would by necessity interpret stress as being unpredictable (or at least subject to a great number of lexical exceptions).7

7 In the hypothetical Gujarati-based example Kiparsky (2004) constructs, this is in fact rather trivially obvious. Given that [ə] already exists in the language, the [a] > [ə] merger would result in the language having two kinds of [ə], one stress-repelling and one stress-attracting. An antepenult or final [ə] would thus attract stress away from a mid or high penult vowel in some...
Citing the same thought experiment on vowel change in a Gujarati-like sonority-driven stress system, de Lacy & Kingston (2006) take a much stronger position. Whereas Kiparsky (2004) expects the sound change to necessitate covert reanalysis and lexicalization of the stress system on the part of language learners, de Lacy & Kingston go one step further and assert that it will—simultaneously and by blind necessity—trigger overt change in the stress patterns of the relevant words (i.e. ones where a stress-attracting [a] is turning into [ə]). They claim that in a language with sonority-driven stress, a sound change like [a] > [ə] ‘will necessarily alter the stress in words that have undergone the change’ (de Lacy & Kingston 2006:7–8). In other words, in a word like [tádʒetəɾ], ‘the *[a] > ə* change must be simultaneous with the change in stress position to [tədʒetəɾ] — there is no stage in the language’s history which would have [tódʒetəɾ]’ (ibid.). It is difficult to see how this position could be integrated with current theories of sound change such as those of Ohala (see §3) or Blevins (see §5), not to mention exemplar models or usage-based models like that of Bybee (2001).

The more moderate interpretation, by which sound patterns that are synchronically ‘impossible’ (i.e. violate synchronic universals) may arise but will simply not be captured as such by learners, is not implausible in principle. 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/:/θ/ (and similarly for obstruents at other places of articulation). Perhaps unsurprisingly, a steady stream of borrowings over the past few centuries have gradually been filling the distributional gap (/ratar/ ‘radar’, /lɛkou/ ‘lego blocks’). This could be taken as evidence that the postvocalic ‘neutralization’ of Tʰ : T in favor of aspirated stops was never an aspect of the Northern Icelandic phonological system in the first place, even at the earlier stage when the distribution facing language learners—i.e. the (supposedly ‘accidental’) gap—would have been completely exceptionless. Variant pronunciations in some items suggest that aspiration has occasionally been imposed, however (e.g. /tʰupa/ ~ /tʰupa/ ‘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

words but not others. Similarly, a mid or high vowel in the antepenult would attract stress away from a penult [ə] in some words but not others. This unfortunate flaw in the particular example chosen does not detract from what is almost certainly the main point of Kiparsky’s argument. We might as well imagine a sonority-driven stress system where the relevant scale is [a] > [e, o] > [i, u], with no [ə] vowel in the language; a sound change [a] > [ə] would indeed turn this into a system in which [ə] attracts stress over and beyond more sonorous vowels like [o] or [e]. Presumably Kiparsky would expect this, too, to result in lexicalization of stress in the hands of subsequent generations of learners, in spite of the fact that stress placement would still be completely predictable, observationally speaking.
is that it is extremely hard to falsify in principle. In practice it is all too easy to explain away apparent counterexamples (alternations or distributional patterns which violate some alleged synchronic universal) as being lexicalized, morphologized, or in some other way 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 there exists a generative phonological module of grammar brings with it certain complications, particularly if this module is assumed to manipulate symbolic elements which are transduced into the phonetic domain by some ‘phonetic implementation’ function. Ambiguities about whether a certain descriptive generalization constitutes an aspect of the knowledge system (in the generative sense)—or, if it does, ambiguities about how that generalization is 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 underdeveloped), 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 model of grammar, 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; interestingly, in a later version Blevins 2006a:124–125 downplays the synchrony/diachrony divide by substituting ‘phonological’ vs. ‘extra-phonological’ for ‘synchronic’ vs. ‘diachronic’). This general idea is of course far from new (see §1), nor is it confined to phonology. In the domain of morphosyntax, for example,
diachronic explanations of the typology of so-called ‘split ergative’ systems (Anderson 1988; Garrett 1990) are widely accepted (e.g., Lightfoot 1999:141; but see Kiparsky 2004 for counterarguments). However, the territorial dispute between diachronic and synchronic/UG-based explanations has intensified sharply in the context of Optimality Theory in phonology, in which functionally ‘grounded’ constraints are standardly proposed and attributed (usually) to an innate UG, as described in §2 above.

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, Blevins and others are in a very fundamental sense reductionist: the explanandum (some recurrent sound pattern or typological generalization) is accounted for in terms of an explanans that is 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 to be 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’ and when it does not.

Another claim made by de Lacy (2006) is that a theory 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 theory 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 constraint like the *ND constraint suggested by Hyman (2001; see above) is no less an instance of a Markedness constraint than its functionally grounded near-opposite *N̥. Likewise, the notion of Markedness constraints does not in itself 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’ from synchronic grammars of ranked and violable output constraints as such, but rather a rejection of the general strategy of accounting for 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 a clear phonetic basis, whereas the comparatively rare occurrence of certain other conceivable sound patterns (sometimes to the point of their being completely unattested) is attributed to the relative inaccessibility of these patterns via such recurrent types of sound changes. As rightly noted by some critics of diachronic approaches, standard generative models of phonology, such as Optimality Theory, have nothing to say about relative typological frequencies as such, 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 a set of learning biases, which do provide a means of dealing with typological asymmetries that are gradient rather than all-or-nothing.

In a series of recent studies, 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. Comparing patterns that relate tone to tone (e.g., of vowels in adjacent syllables) against ones that relate tone to laryngeal features (e.g., the voicing or voicelessness 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): even though phonetic vowel height is affected more substantially by the voicing or voicelessness of an adjacent consonant than it is by the height of a vowel in a neighboring syllable, phonological vowel height harmony is much more common than phonological vowel raising/lowering conditioned by the voicing of adjacent consonants. On the assumption that the relative magnitude of a phonetic effect ought to determine (or at least be positively correlated with) 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 interactions (tone-tone, height-height).

This novel approach to the problem of phonologization is certainly very promising, but while Moreton’s underphonologization argument appears compelling at first glance, it rests on certain assumptions which may be overly simplistic. Misperception (miscalculation, misparsing) is the key process in listener-based models of sound change, in that the listener either overcorrects or undercorrects for a contextual effect (be it real or apparent) in the acoustic signal. However, it is far from obvious that a direct parallel can be drawn between some raw acoustic-auditory measure of the ‘magnitude’ of an interaction on the one hand and how that interaction is going to be dealt with by the listener in the perception/parsing process on the other. The kinds of phenomena that Moreton deals with would all fall under Ohala’s hypocorrection rubric, whereby coarticulatory compensation (Beddor & Krakow 1999; Beddor et al. 2002) is underapplied, such that a coarticulatory effect is misinterpreted as an intrinsic property of the affected segment. Ohala (1994) has suggested that listeners may be more attuned to—and hence more likely to compensate successfully for—the influence of stronger coarticulation triggers than that of weaker triggers. For example, this would explain why in most front/back vowel harmony systems, it is precisely those vowels which ought to be the strongest inducers of coarticulatory fronting on neighboring vowels (namely /i, e/) that are instead ‘neutral’, failing to trigger any phonological fronting at all. If Ohala’s (1994) theory of asymmetric coarticulatory compensation has general validity, this would predicts that interactions with greater ‘magnitude’ (more robust phonetic precursors, in Moreton’s sense) ought in fact to be less likely to become phonologized. Needless to say, this is exactly what the typological data show; the case for ‘underphonologization’ may thus turn out to be illusory.

Such caveats aside, a second and no less important aspect of Moreton’s work is his demonstration that artificial language-learning experiments yield results which do mirror the typological distribution closely: cross-dimension patterns (height-voicing interaction) are learned far less reliably than same-dimension patterns are (Moreton 2007). In this particular case, the factor which governs 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 supposed underphonologization of cross-dimension interactions in terms which go beyond superficial measures of formal simplicity. He proposes an explicit Bayesian learning algorithm (set within an Optimality Theory framework) in which the language learner is biased against adding a new constraint—that is, inventing or constructing

---

8 Interestingly, the more easily learned same-dimension patterns include not only vowel height harmony but also (long-distance) consonant voicing harmony, which is typologically quite rare in comparison and may be largely dependent on complex diachronic scenarios for its emergence (Hansson 2004).
it on the basis of input data—if that constraint interacts with a large number of other constraints that are already present in the grammar.

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] tended to generalize the [k] → [tʃ] change to [ki] contexts as well, whereas subjects who learned [ki] → [tʃi] did not generalize this to [ke] contexts. Wilson (2006) argues for a conception of markedness and phonetic 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 obstruent-obstruent clusters like /bd/ than of 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 via Latin to Modern French, Martin (2006, 2007) 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, they also appear to be favored in the formation of new words through morphological derivation, as well as in borrowing. For example, even though /d/-initial words outnumbered /b/-initial ones both in the Latin lexicon and in that of 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 by no means incompatible with diachronic-functional approaches. For example, if such biases do exist, they would simply be one of the many potential sources of ‘external error’ in evolutionary 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]
to take place before [i] than before [e]), this does most certainly complicate the task of teasing apart the true explanatory factors which underlie specific typological generalizations. More research in this area is clearly needed in order to shed further light on this contentious issue.

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 more active than it has ever been in the history of modern phonological theory. This has partly been brought on by the formulation of elaborate and explicit theories of sound change. In addition, our knowledge of the full range of typological variation in the sound systems of the world's languages has become much more principled and more detailed—helped along in large part by the inherently typological orientation of the Optimality Theory enterprise—and this 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 methodologies and approaches 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 generative phonological theory will be shaped by these developments over the next decade or two.

Acknowledgements

I would like to thank Andrew Garrett, Cathie Ringen and two anonymous reviewers for helpful comments on an earlier draft of this article; I am particularly grateful to Andrew for inspiring my thinking about language change, diachronic explanations and the synchrony-diachrony relationship over many years. Special thanks are also due another of my mentors, John Ohala, whose work on the phonetics of sound change fascinated me long before I became a Berkeley graduate student. Finally, I am indebted to Doug Pulleyblank, Jason Brown, Shane Jobber and participants in a graduate course on the phonetics-phonology interface at the University of British Columbia for stimulating discussion of many of the issues which are covered here. The work underlying this article was partly supported by Social Sciences and Humanities Research Council of Canada grant 410-2004-0710.

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.


---

**Short Biography**

Gunnar Ólafur Hansson is Assistant Professor in the Department of Linguistics at the University of British Columbia and Faculty Associate at the Peter Wall Institute for Advanced Studies. He previously taught at the University of Chicago. His primary research area is theoretical phonology, the interfaces of phonology with phonetics, morphology and the lexicon, and the interplay of diachronic and synchronic factors in explaining the properties of linguistic systems and cross-linguistic typologies. Much of his current research revolves around issues of locality in segmental phonology, exploring the nature and sources of non-local (long-distance) interactions in phonological harmony systems (especially consonant harmony). Other research topics include the problem of derivationally opaque sound patterns, and various issues relating to morphological paradigms and their manifestation: inflectional classes, allomorphy, syncretism patterns and paradigm gaps. He has published in journals like *Phonology* and *Linguistic Inquiry* and in various edited volumes and conference proceedings. He holds an M.A. from the University of Iceland and a Ph.D. from the University of California, Berkeley.

**Correspondence address:** Gunnar Ólafur Hansson, Department of Linguistics, University of British Columbia, Totem Field Studios, 2613 West Mall, Vancouver, B.C., Canada V6T 1Z4. Email: ghansson@interchange.ubc.ca