

21 Diachronic Phonology

Ricardo Bermúdez-Otero

University of Manchester

21.1 Introduction*

As the title of this part of the volume indicates, the study of *sound change*¹ compels us to think hard about the relationship between phonological structure and what lies beyond: the physics of sound, the physiology of speech, the social and cultural context of communication. Yet, for that very reason, diachrony lies at the heart of current phonological debates. In particular, historical questions are crucial to the renewed controversy between formalist and functionalist approaches to phonology, which respond in different ways to a fundamental fact: phonological structure is moulded by external forces through change, but also imposes constraints on the possible courses of change (see Gordon [ch.3] and Kingston [ch.17]).

One of the basic challenges for diachronic phonology is the problem of *innovation*: how does a phonological variant that has never existed previously in a speech community first come into being? Here, it is commonly agreed that the potential for innovation leading to sound change arises whenever speaker and listener fail to solve the *coordination problem*² posed by speech: the speaker must produce a phonetic stimulus that enables the listener to recover the intended phonological representation; the listener must decide which properties of the incoming stimulus are intended by the speaker as signal, and which properties are accidental noise; neither participant can read the other's mind. The innovation mechanisms proposed by Ohala, *hypocorrection* and *hypercorrection*, both involve failures of coordination: the listener does not parse the stimulus in the way that the speaker intended (see e.g. Ohala 1989, Alderete & Frisch [16.3], Kingston [17.3.3]). Beyond this point, however, disagreement rages over important questions:

- (i) What are the relative rôles of the speaker and the listener in bringing about a coordination failure?

* I am grateful to Paul de Lacy, Paula Fikkert, Randall Gess, Larry Hyman, Donka Minkova, and James Scobbie for their comments and suggestions.

¹ Phonological change is traditionally divided into *sound change*, initiated by phonetic causes, and *analogy*, driven by morphological factors. This chapter concentrates on the former. Bermúdez-Otero (forthcoming a) addresses the study of analogy in Optimality Theory (OT, Prince & Smolensky 1993); Bermúdez-Otero & Hogg (2003:§3) provide an illustrative case study. In §21.3.2 below I discuss the rôle of sound change and analogy in the life cycle of sound patterns.

² For general discussion of the coordination problem in linguistic communication, see Croft (2000:95ff).

- (ii) Do the crucial coordination failures that lead to innovation happen when the listener is a child acquiring language or when the listener is an adult?
- (iii) To what extent are innovations driven and controlled by *bottom-up* factors (e.g. phonetic effects) and *top-down* biases (e.g. phonological knowledge)?

Question (iii), in particular, has figured prominently in the debate concerning the nature and origins of phonological markedness (see Gordon [3.5], Rice [ch.4]): some phonologists argue that markedness is a mere epiphenomenon of sound changes actuated by bottom-up factors; others claim that knowledge of markedness imposes top-down constraints on innovation (Bermúdez-Otero & Börjars 2006; Bermúdez-Otero forthcoming a).

This chapter will largely focus on another challenge: describing and explaining the time course of sound change. Is sound change implemented gradually or abruptly, and why? As we shall see, this question too has profound implications for the nature of phonological representations and the architecture of phonological grammar.

21.2 The implementation problem: how gradual is phonological change?

In a pretheoretical sense, all phonological change is gradual: developments such as the raising of /ɑ:/ to /ɔ:/ in southern dialects of Middle English – and, *a fortiori*, large-scale upheavals like the Great Vowel Shift – do not take place overnight. However, this obvious fact does not imply that phonological change advances gradually in all dimensions.

One must first distinguish between graduality in *implementation* and graduality in *propagation*. For example, Prehistoric Latin is reconstructed as having left-dominant word stress: the first syllable of the word was the most prominent. Later, Classical Latin developed a right-dominant pattern whereby primary stress was assigned to the penultimate syllable if heavy, otherwise to the antepenult: e.g. prehistoric *má.le.fi.ci.um* > classical *mà.le.fī.cī.um* ‘bad deed’ (Allen 1973: 189-190). Imagine that, while Latin was undergoing this change, every individual speaker fell into just one of two groups: one pronouncing all words invariably with a left-strong contour, the other pronouncing all words invariably with a right-strong contour. Had that been the case, the shift from left to right dominance would have been implemented abruptly: there would have been graduality only in its propagation, as the proportion of speakers with right-dominant stress increased in the community over time. As we shall see presently, however, implementation is never completely abrupt; rather, every phonological change is implemented gradually in one or more of the following dimensions: sociolinguistic, phonetic, and lexical.

Sociolinguistic research indicates that all phonological changes involve a transitional phase of variation (Anttila [ch.22]). Therefore, one can confidently assume that, while Latin was undergoing its shift from left to right dominance, an individual speaker might pronounce the same word sometimes with primary stress on the initial syllable, sometimes with primary stress on the penult or antepenult: e.g. *má.le.fi.ci.um* ~ *mà.le.fī.cī.um*. The relative frequency with which speakers used left-strong or right-strong stress probably reflected external factors such as sex, age, social status, and so

forth. In this sociolinguistic dimension, phonological change typically advances through generational increments: successive generations of speakers use the innovative variant with increasing frequency (Labov 1994:84; 2001:Part D). The omnipresence of variation during change in progress is one of the reasons why quantitative techniques are indispensable in research into the problem of implementation.

A change is said to be phonetically gradual – or *gradient* – if it involves a continuous shift along one or more dimensions in phonetic space, such as the frequency of the first formant of a vowel as measured in hertz. In contrast, a change is phonetically abrupt – or *categorical* – if it involves the substitution of one discrete phonological category for another: e.g. replacing the feature [–high] with [+high] (see Harris [6.2.1]). Deciding whether the pattern created by a change is gradient or categorical often requires careful instrumental analysis, as well as a global understanding of the phonology-phonetics interface in the language in question (Myers 2000). Indeed, laboratory research has in recent times redressed the balance between gradient and categorical rules in phonology (§21.3.1). Languages have been shown to vary with respect to the phonetic realization of phonological categories down to the finest detail: for example, contrary to the assumptions of *SPE* (Chomsky & Halle 1969: 295), patterns of coarticulation are not mechanical and universal, but cognitive and acquired (Keating 1988: 287-288; Pierrehumbert *et al.* 2000: 285-286). In addition, many phenomena previously thought to be categorical have proved to be gradient (Myers 2000: 257).

The ongoing lengthening and raising of the reflexes of Middle English short /a/ in contemporary American English (henceforth ‘æ-tensing’) provides a striking illustration of the contrast between gradient and categorical implementation (Labov 1981, 1994). In the northern dialect area comprising cities such as Albany, Rochester, Buffalo, Detroit, and Chicago, æ-tensing is phonetically gradual: the allophones of /æ/ form an unbroken phonetic continuum from the highest and most peripheral (e.g. in *aunt*) to the lowest and less peripheral (e.g. in *black*); the degree of tensing displayed by each allophone is exquisitely sensitive to a broad range of properties of its phonetic environment (Labov 1994:456-459; see also Gordon 2001:124-140). In Mid-Atlantic cities such as New York, Philadelphia, and Baltimore, in contrast, æ-tensing is phonetically abrupt: lax /æ/ and tense /æ:/ have widely separated phonetic targets, namely low [æ] vs. mid-high [e:⁹], and their tokens occupy discrete, largely nonoverlapping regions in phonetic space (see e.g. Labov 1989: 7-11).

As a first approximation, lexically abrupt – or *regular* – implementation can be defined as follows: a change is regular if it applies at the same time to all words that are identical with respect to the relevant phonological, morphological, and syntactic conditions. In contrast, a change is lexically gradual – or *diffusing* – if it affects certain words earlier than others with an equivalent phonological and morphosyntactic makeup, i.e. if lexical identity plays an irreducible rôle in controlling the advance of the change. When applying this definition in practice, one must take account of sociolinguistic variation: lexical diffusion can manifest itself through a difference in the relative frequency with which two words display the innovative variant, as long as this difference is not determined by phonological, morphological, or syntactic conditions, or by sociolinguistic factors (e.g. sex, age, social status, style, register, etc.). Accordingly, establishing whether a particular change is regular or diffusing often requires large datasets and powerful statistical methods (e.g. Labov 1994:ch.16).

In this connection, a particularly effective way of controlling for unknown phonological factors is to focus on the behaviour of homophones: e.g. the English words /tu:/ ‘two’ and /tu:/ ‘too’. When two initially homophonous words cease to be phonologically identical by undergoing different processes of change, we have strong evidence for lexical diffusion (Chen 1972:§6). The Chao-zhou dialect of Chinese provides a notable instance of this phenomenon, known as a *homonym split*: in Chao-zhou, twelve pairs of homophonous Middle Chinese words with tone III have become split between the modern tones 2b and 3b (Cheng & Wang 1973; but see §21.4.2 below). In contrast, Labov (1994: 460-465) shows that, in Philadelphia English, homophones such as *two* and *too* undergo the fronting of /u:/ to [ʊ:^u] at exactly the same rate, and on this basis he argues that the change is regular.

Interestingly, Labov also shows that *æ*-tensing is lexically abrupt in the northern cities of the United States, but lexically gradual in the Mid-Atlantic region. In Philadelphia, for example, there is incipient tensing of *æ* before /d/: in particular, *æ* has become tense in the affective adjectives *mad*, *bad*, and *glad*. This is an innovation with respect to the tensing pattern found in Early Modern English: cf. British Received Pronunciation, which has /mæd/, /bæd/, and /glæd/, rather than */mɑ:d/, */bɑ:d/, or */glɑ:d/. Nonetheless, the tensing of *æ* before /d/ in Philadelphia remains lexically idiosyncratic: most noticeably, the vowel remains lax in the affective adjective *sad* (Labov 1994: 429-437). The contrast between *gl[e:^ɨ]d* and *s[æ]d* is particularly striking, as elsewhere tensing is strongly disfavoured after an obstruent+liquid cluster (Labov 1994: 433, 458-459).

In the 1980s and 1990s, the work of William Labov and Paul Kiparsky brought about a convergence of empirical results and theoretical perspectives on the implementation problem. Labov’s (1981, 1994) empirical findings confirmed the existence of two long-recognized mechanisms of phonological change: *Neogrammarian change* (Osthoff & Brugmann 1878) and *classical lexical diffusion* (Wang 1969). The former (exemplified by *æ*-tensing in the northern cities) is regular but gradient; the latter (instantiated by *æ*-tensing in the Mid-Atlantic states) is categorical but diffusing.

(1) *The implementation of phonological change: the received view*

		Dimensions	
		Phonetic	Lexical
Modes	<i>Neogrammarian sound change</i>	gradual	abrupt
	<i>Classical lexical diffusion</i>	abrupt	gradual

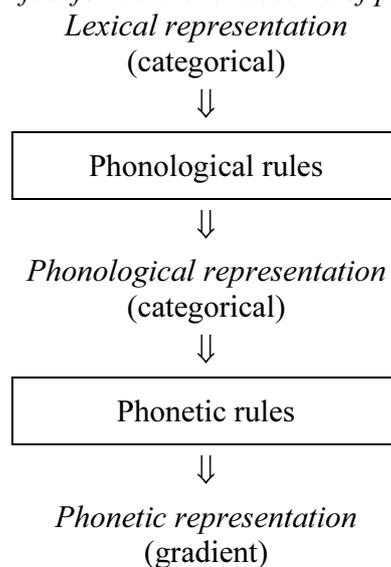
Kiparsky (1988, 1995) then showed how the existence of these two modes of implementation follows from the architecture of grammar in generative theory, particularly in Lexical Phonology. Lately, however, the received view has come under challenge, as the claim that all phonological change is both lexically and phonetically gradual (Bybee 1998, 2001) gains increasing currency. The ensuing debate bears directly on a central issue in phonological theory: whether or not lexical representations contain gradient phonetic detail.

21.3 The view from generative phonology

21.3.1 Modular feedforward models: phonological rules vs phonetic rules

Structuralist and generative theories of phonology assume a fundamental distinction between phonological and phonetic rules,³ where the term *rule* is to be understood in its widest sense as meaning ‘symbolic generalization’. Such symbolic generalizations may be instantiated by various devices, including the input-driven transformations of *SPE* and the output-oriented constraints of *OT*; but in the discussion that follows I will in general not be concerned with choosing among them. The distinction between phonological and phonetic rules is typically embedded in the general grammatical architecture shown in (2); see Kingston [17.4.3]. Following Pierrehumbert (2002), I shall describe all versions of (2) as ‘modular feedforward models’.⁴ This grammatical architecture is equally compatible with the view that modules are innate (e.g. Fodor 1985) and with the idea that modularity emerges during the child’s cognitive development (e.g. Karmiloff-Smith 1994).

(2) The classical modular feedforward architecture of phonology



Modular feedforward models of phonology rest on two key assumptions (see the discussion in Pierrehumbert 2002:101-102):

³ Phonetic rules are often known as ‘rules of phonetic implementation’. In the current discussion, therefore, the word *implementation* will occur in two separate senses: ‘historical implementation of phonological changes’, and ‘phonetic implementation of phonological categories’.

⁴ The currently prevalent version of *OT* is a modular feedforward model, since it assumes that phonological constraints generate the input to phonetic implementation. Parallelism is stipulated to hold within the phonological module, but not in the relationship between the phonological and phonetic modules.

- (3) (a) *Lexical and phonological discreteness*
 In lexical and phonological representations, attributes have discrete values.
- (b) *Modularity*
 Phonetic rules cannot refer directly to lexical representations.

An example of (3a) is the common postulate that, in the phonological module, distinctive feature specifications are maximally ternary: e.g. a segment can at most be [+voice], [−voice], or underspecified for voicing. Here, however, the term *attribute* is intended in a wide sense as including any relevant phonological property: e.g. the presence or absence of an association line between two nodes. Thus, principle (3a) prevents lexical and phonological representations from encoding fine phonetic detail, which would require gradient attribute values. Note, however, that (3a) makes no claim as to whether lexical representations should be minimal (cf. Bybee 2001:§3.4.2-§3.4.3). Some versions of the modular feedforward architecture assume that lexical representations do not contain any predictable information (e.g. Kiparsky 1982, 1995; Archangeli 1984), but others reject this claim (e.g. Prince & Smolensky 1993/2004, Steriade 1995, Bermúdez-Otero & McMahon forthcoming). Principle (3a) allows allophonic information to be stored in lexical representations as long as this information is categorical rather than gradient.

The principles in (3) give expression to fundamental phonological assumptions. One is the idea that, on the phonological no less than on the syntactic side, linguistic expressions arise from the combination of a small inventory of elements: see Martinet's (1960) notion of the double articulation of language. Given (3a), phonological representations cannot behave as holistic articulatory or auditory patterns because they do not contain continuous phonetic information, but are rather composed of discrete units. Phonologists have also adduced various types of empirical evidence in support of (3). Of particular relevance here are the arguments from diachrony.⁵ In line with the received view of implementation outlined in (1), the principles in (3) account for the regularity of gradient changes (Neogrammarian change) and the categorical nature of diffusing changes (classical lexical diffusion):

- (i) By (3a), phonetically gradual change can take place only through the alteration of the phonetic rules that assign realizations to phonological categories. But, by (3b), any such alteration must be free of lexical conditioning. This is the key insight behind Bloomfield's (1933:351) slogan 'Phonemes change'.
- (ii) Diffusing change involves the alteration of the lexical representations where lexical information is stored. By (3a), however, such alterations must be categorical.

In fact, given all the logically possible interactions between the phonetic and lexical dimensions, the architecture in (2) predicts the existence of not two but three modes of implementation for phonological change:

⁵ There is also, for example, a long tradition of psycholinguistic research that relies on the classical modular feedforward architecture for the explanation of speech errors (e.g. Fromkin 1971).

(4) *Modes of implementation predicted by the classical architecture*

Mode of implementation		Possible?	Innovation in what component of grammar?
<i>phonetic dimension</i>	<i>lexical dimension</i>		
abrupt	gradual	Yes	lexical representations
abrupt	abrupt	Yes	phonological rules
gradual	abrupt	Yes	phonetic rules
gradual	gradual	No	

To my knowledge, the prediction that regular categorical change exists has rarely, if ever, been explicitly discussed in the literature. As we shall see in §21.3.2, however, there are powerful arguments in its favour.

21.3.2 *The life cycle of sound patterns, stabilization, and secondary split*

At least since Baudouin de Courtenay (1895), phonologists have acknowledged that sound patterns evolve historically according to a characteristic *life cycle*. Modular feedforward models – specially Lexical Phonology and Stratal OT— provide a perspicuous interpretation of this observation (Kiparsky 1988; McMahon 2000; Bermúdez-Otero forthcoming a, b):

• *Phase I*

The life cycle starts with *phonologization* (Hyman 1976), which occurs when some physical or physiological phenomenon gives rise to a new cognitively controlled pattern of phonetic implementation through a coordination failure (see §21.1). This development involves the addition of a new phonetic rule to the grammar and manifests itself as Neogrammarian sound change (i.e. regular gradient change: see (1) above).

• *Phase II*

Subsequently, the new gradient sound pattern may become categorical. In the modular feedforward architecture in (2), such a change would involve the *restructuring* of the phonological representations that provide the input to phonetic implementation, with the concomitant development of a new phonological counterpart for the original phonetic rule. As we shall see below, this step in the life cycle corresponds to the process of ‘stabilization’ discussed in Hayes & Steriade (2004:§5.6), and is implicated in the rise of so-called ‘quasi-phonemes’ that precedes secondary split (Janda 1999).

• *Phase III*

Reanalysis can also cause categorical patterns to change. Over time, phonological rules typically become sensitive to morphosyntactic structure, often with a reduction in their domain of application (Dressler 1985:149). In models such as Lexical Phonology and Stratal OT, these changes involve the ascent of

phonological rules from the phrase level through the word level into the stem level: see Bermúdez-Otero & Hogg (2003) and Bermúdez-Otero (forthcoming a, b) for discussion. Phonological rules may also develop lexical exceptions: cf. the lexical split of short /a/ in southern British English (§21.4.2 below).

- *Phase IV*

At the end of their life cycle, sound patterns may cease to be phonologically controlled. Thus, a phonological rule may be replaced by a morphological operation (*morphologization*), or may disappear altogether, leaving an idiosyncratic residue in lexical representations (*lexicalization*). See Anderson (1988) for examples and discussion.

Let us now focus on Phase II: the emergence of a phonological rule from a phonetic one. By definition, changes of this sort are phonetically abrupt, since they have the effect of creating a new distribution of discrete categories in the output of the phonological module. However, the new phonological rule may remain free of lexical idiosyncrasies, for restructuring may fall short of altering the content of lexical representations (cf. Phase III above); in that case, the change whereby the new rule is introduced will be lexically abrupt too. In this sense, the extra mode of implementation predicted in diagram (4), viz. regular categorical change, finds a natural niche in our account of the life cycle of sound patterns. Empirically, however, developments of this type may be difficult to detect: as we saw in §21.2, the distinction between categorical and gradient rules cannot be reliably drawn on the basis of impressionistic data and, indeed, the change may be largely covert. Of relevance here is Karmiloff-Smith's (1994:700) distinction between behavioural change and representational change during the child's cognitive development: 'The same performance (say, correctly producing a particular linguistic form, or managing to balance blocks on a narrow support) can be generated at various ages by very different representations.' In terms of Karmiloff-Smith's cognitive science metatheory, sound patterns reach Phase II in their life cycle by a process of *representational redescription*.

Regular categorical changes involving the rise of a new phonological generalization out of an existing phonetic rule have the effect of *stabilizing* patterns of allophonic variation (cf. Hayes & Steriade 2004:§5.6). As noted in Myers (2000:§6.3), phonetic rules are gradient in respect not only of their effects, but also of the factors that condition them: for example, Sproat & Fujimura (1993) showed that, in the pronunciation of dark [ɫ] in American English, the delay of the coronal C-gesture relative to the dorsal V-gesture increases continuously in proportion to the duration of the syllable rhyme.⁶ The modular feedforward architecture in (2) predicts that, in contrast, phonological rules will not be sensitive to quantitative properties of the phonetic environment, since information about such properties is absent at the phonological level. In Modern

⁶ *Pace* Sproat & Fujimura (1993), however, gradient variation in the phonetic realization of dark [ɫ] is not incompatible with a categorical distinction between light [l] and dark [ɫ]; in fact, the modular feedforward architecture predicts the existence of such a distinction, for in American English dialects the distribution of the light and dark allophones of /l/ is sensitive to morphosyntactic structure (Phase III above; see Hayes 2000: 93). As we shall see below, phonological generalizations typically coexist in the same grammar with the gradient phonetic rules from which they emerge historically.

Japanese, for example, the phonemic opposition between /t/ and /t̥ç/ is neutralized before /i/, where only [t̥ç] occurs: see (5a,b). This rule probably arose historically through the phonologization of a coarticulation effect in [ti] sequences: anticipating the gesture of tongue front raising for [i] narrows the channel for the release of [t], causing the stop burst to become relatively noisy. In present-day Japanese, however, reducing the amount of CV coarticulation fails to restore the contrast between /t/ and /t̥ç/ in the neutralization environment: even in the most careful *hyperspeech* (Lindblom 1990), the realization of /t/ before /i/ remains [t̥ç] (Mitsuhiko Ota, personal communication). The modular feedforward architecture correctly predicts this state of affairs: affrication cannot be blocked by gradient adjustments to gestural timing because it is a categorical rule. This is independently shown by the fact that it has lexical exceptions among loanwords, and so has already progressed to Phase III in its life cycle: see (5c) and cf. Bybee (2001: 53).

(5) *Japanese /t/-affrication (Itô & Mester 1995: 827-828)*

- | | | | | |
|-----|-------------|-----------|-----|-------------------|
| (a) | /ta/ | /t̥ça/ | | |
| | [ta] | [t̥ça] | | |
| | ‘field’ | ‘tea’ | | |
| (b) | /kat-i/ | /kat-e/ | | |
| | [kat̥çi] | [kate] | | |
| | ‘win-INF’ | ‘win-IMP’ | | |
| (c) | [t̥çimɯ] | ‘team’ | but | [ti:N] ‘teenager’ |
| | [t̥çiket:o] | ‘ticket’ | | [pa:ti:] ‘party’ |

Interestingly, the innovative phonological rules created by stabilization do not replace the phonetic rules from which they emerge, but typically coexist with them. English palatalization provides a well-studied example of this coexistence: categorical palatalization in stem-level domains, e.g. *confess* [kən'fɛs] ~ *confession* [kən'fɛʃn], coexists with gradient palatalization by gestural overlap across word boundaries, e.g. *press you* [ˈpɹɛsʃu:] (Zsiga 1995, discussed in Kingston [17.4.3]). For a diachronic illustration, see the discussion of the Scottish Vowel Length Rule and low-level lengthening in McMahon (2000:ch.4, specially §4.5.2).

In sum, stabilization is a process of regular categorical change that creates a new phonological counterpart for an existing phonetic rule. Understood in this sense, stabilization is a prerequisite for *secondary split*. This structuralist term designates a historical development whereby the destruction of the conditioning environment of an allophonic rule gives rise to a new phonemic opposition (see Fox 1995:§3.2). The Sanskrit Law of Palatals provides a justly celebrated example (see Fox 1995:27-29):

(6) *The Law of Palatals*

- | | | | | | |
|--------------------------------------|-------|-------|------------|-------|-------|
| (a) <i>Proto-Indo-Iranian</i> | *-ki- | *-ke- | *-ka- | *-ko- | *-ku- |
| (b) <i>Palatalization</i> | *-ci- | *-ce- | *-ka- | *-ko- | *-ku- |
| (c) <i>Lowering of /e, o/ to [a]</i> | *-ci- | *-ca- | *-ka- | *-ka- | *-ku- |
| (d) <i>Sanskrit distribution</i> | -ci- | | - { k } a- | | -ku- |
| | | | - { c } a- | | |

At the synchronic stage represented by (6b), [k] and [c] are in complementary distribution: the language has a single phoneme /k/, allophonically realized as [c] before nonlow front vowels. At a later point, /e/ and /o/ undergo lowering to [a]. This development removes the trigger of palatalization in the reflexes of Proto-Indo-Iranian *-ke-. In consequence, the distinction between [k] and [c] becomes phonemic in Sanskrit, as the two phones are in contrastive distribution before [a].

It is crucial for this development that, at the synchronic stage represented by (6c), the rule of lowering should *counterbleed* the rule of palatalization: i.e. lowering must apply to the output of palatalization, removing the cause of palatalization without reversing its effect; see (7a). This synchronic interaction between the two rules is an instance of *opacity* in the sense defined by Kiparsky (1971) (see McCarthy [ch.5]). The counterbleeding relationship between palatalization and lowering caused the restructuring of lexical representations, with the opaque string [-ca-] being reanalysed as underlying; see (7b).

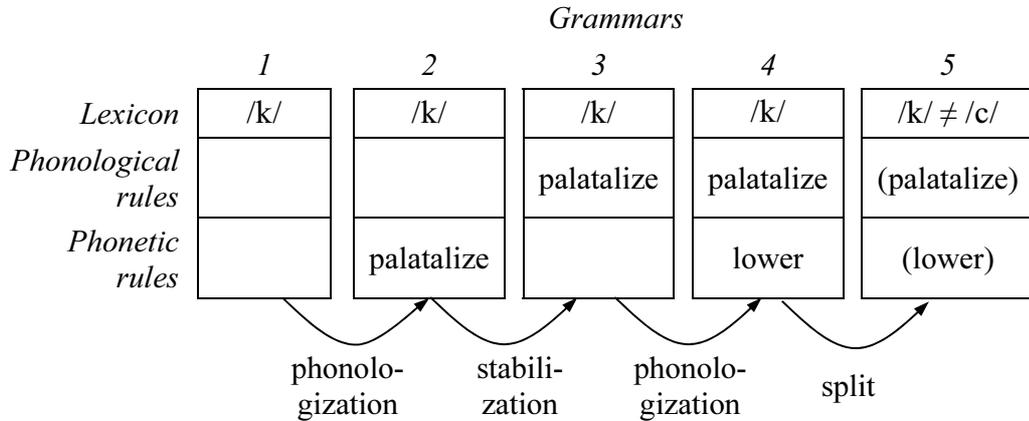
(7)	(a) <i>Opaque derivation</i>	/-ke-/		/-ka/		/-ko-/
	Palatalization	-ce-	}	counterbleeding	-	-
	Lowering	-ca-	}		-	-ka-
	(b) <i>Reanalysis</i>	/-ca-/				/-ka-/

If, upon entering the grammar, lowering had interacted transparently with palatalization, [k] and [c] would have remained in complementary distribution:

(8)	<i>No secondary split without synchronic opacity</i>	/-ke-/		/-ki-/	/-ka-/	/-ko-/
	Lowering	-ka-	}	bleeding	-	-ka-
	Palatalization	-	}		-ci-	-

This raises an obvious question: why was the interaction opaque rather than transparent? Assuming the ordinary life cycle of sound patterns as implemented in the classical modular feedforward architecture provides a straightforward solution. Presumably, both palatalization and lowering first came into being as phonetic rules. However, when lowering entered the phonetic implementation module as a gradient pattern, palatalization had already undergone stabilization and become a categorical phonological rule.

(9) *Stabilization precedes secondary split*



If this analysis is on the right track, the original phoneme /k/ could split only after its allophones [k] and [c] had become discrete phonological categories through the restructuring of the input to the phonetic module. This proposal provides a formal interpretation of the claim that only *quasi-phonemes* can become phonemicized through the loss of their conditioning context (see Kiparsky 1995:657; Janda 1999; and references therein): discrete allophones generated by categorical phonological rules are quasi-phonemes in the intended sense; nondiscrete allophones created by gradient phonetic rules are not.

In nonmodular theories of the phonology-phonetics interface, in contrast, a predictable allophone can remain after the loss of its conditioning environment only if it is already present in lexical representation (Bybee 2001:§3.6). However, this approach misses the crucial rôle of opacity in triggering lexical restructuring: see (7). Modular feedforward models capture the right sequence of cause and effect: in the Sanskrit case, the stabilization of palatalization enabled it to interact opaquely with lowering; this opacity, in turn, prompted lexical restructuring, with the attendant phonemic split.

Note, in addition, that an allophonic pattern may undergo stabilization without necessarily becoming word-bound (i.e. ‘lexical’ in the sense of Lexical Phonology). The distinction between word-bound rules and phrasal rules does not coincide with the distinction between categorical and gradient processes, for a rule may apply across word boundaries without being gradient. In the case of English palatalization (Zsiga 1995), for example, the categorical neutralizing version of the process happens to apply only in stem-level domains, whereas palatalization across word boundaries is the result of gradient coarticulation (see above). In Sardinian, however, assimilatory external sandhi is categorical (Ladd & Scobbie 2003).

21.3.3 *The mechanism of classical lexical diffusion*

In modular feedforward models of phonology, classical lexical diffusion is implemented through category substitution in lexical representations: see (4). Consider, for example, the incipient tensing of *æ* before /d/ in Philadelphia English (§21.2). The surface contrast between [sæd, dæd, læd] and [me:⁰d, be:⁰d, gle:⁰d] shows that there is a lexical

opposition between /æ/ and /æ:/.⁷ Thus, the tensing of *æ* in *mad*, *bad*, and *glad* involved the replacement of /æ/ with /æ:/ in the lexical representation of each of the affected items:

(10)	<i>Lexical diffusion by category substitution in lexical representations</i>						
		‘mad’	‘bad’	‘glad’	‘sad’	‘dad’	‘lad’
(a)	(Early) Modern English	/mæd/	/bæd/	/glæd/	/sæd/	/dæd/	/læd/
(b)	Present-day Philadelphia	/mæ:d/	/bæ:d/	/glæ:d/	/sæd/	/dæd/	/læd/

In this section we shall see that, in addition to providing an implementation mechanism for classical lexical diffusion, modular theories of phonology can also make a partial contribution to our understanding of the causes of diffusing innovations. Kiparsky (1988, 1995) states the key ideas. Classical lexical diffusion is driven by a combination of top-down and bottom-up effects (see §21.1). Phonological rules introduce a language-specific top-down bias in the learner’s expectations regarding the distribution of contrastive features in the lexicon: in particular, the rules designate certain feature values as marked in certain contexts. Under pressure from performance (bottom-up) factors, feature values designated as marked become particularly vulnerable to misperception and therefore misacquisition.

Several observations suggest that, despite their lexical irregularity, diffusing changes are under some sort of phonological control. First, feature substitution in lexical representations is not random, but takes place under fairly well-defined phonological and morphological conditions: in Philadelphia English, for example, *æ* is prone to tensing when followed tautosyllabically within a stem-level domain by an anterior nasal, a voiceless anterior fricative, or (incipiently) /d/.⁸ Secondly, the conditioning factors of diffusing changes are specific to particular languages or dialects: in New York City, for example, the set of consonants that trigger *æ*-tensing is much larger than in Philadelphia (Labov 1994: 430, 520).

(11)	<i>Triggers of lexically gradual æ-tensing in the Mid-Atlantic states</i>				
		<i>New York City</i>		<i>Philadelphia</i>	
(a)	voiced stops	b	d	dʒ	g
(b)	voiceless fricatives	f	θ	s	ʃ
(c)	nasals	m	n		

Thirdly, diffusing changes appear in general not to be phonetically unnatural: indeed, in the case of *æ*-tensing we see the same phenomenon advancing by categorical diffusion in the Mid-Atlantic states and operating as a gradient phonetic rule in the Northern Cities Shift (§21.2). All these arguments indicate that, except for its diffusing character,

⁷ The diffusing substitution of /æ:/ for /æ/ in lexical representations should not be conflated with the raising of lexical /æ:/ to phonetic [e:°]. Labov (1994:§16.5) shows that the raising of /æ:/ to [e:°] is a regular gradient process of phonetic implementation.

⁸ Assigning *æ*-tensing to the stem level accounts for its overapplication in words like *glassy* [gle:°.si]. The root vowel and the following fricative are tautosyllabic at the stem level, before the addition of the suffix -y triggers resyllabification: [Word Level [Stem Level glæ:s] i]. See Kiparsky (1988: 400).

Philadelphia *æ*-tensing produces similar effects to an obligatory phonological rule that, having emerged through the stabilization of a gradient phonetic pattern, had ascended to the stem level, with a concomitant reduction in its domain of application (Phase III in the life cycle outlined in §21.3.2). Kiparsky’s crucial insight was to realize that, if Philadelphia *æ*-tensing behaved so much like a stem-level phonological rule, it was because such a rule was indeed involved. In Kiparsky (1988, 1995), he used Radical Underspecification Theory to formalize his approach; but, as observed in Golsmith (1995a:17), this particular technology is not essential. Bermúdez-Otero (1998:§3.4) indicates how the same idea can be expressed in terms of OT.

Consider, for example, the phonemic opposition between /æ/ and /æ:/ before a tautosyllabic /f/ in Philadelphia English: e.g. *Afghan* [æf.gən]⁹ vs *after* [eːf.təɪ] (Labov 1994:507). Maintaining this contrast requires that the faithfulness constraint IDENT-length should outrank both the context-sensitive markedness constraint *ǣf_σ] and the context-free markedness constraint *ǣ.¹⁰

(12)

Philadelphia		IDENT-length	*ǣf _σ]	*ǣ
/æfgən/	[æ:f.gən]	*!		*
	↗ [æf.gən]		*	
/æ:ftəɪ/	↗ [æ:f.təɪ]			*
	[æf.təɪ]	*!	*	

In addition, assume that in this dialect *ǣf_σ], though dominated by IDENT-length, is ranked higher than *ǣ. The effect of this ordering of the constraints will be to designate the input string /-æfC-/ as marked: although [æf.gən] is the optimal output for input /æfgən/, lexical representations containing long /æ:/ in the same environment nonetheless allow for input-output mappings with a better constraint profile.

(13) *Input /-æfC-/ is marked relative to /-æ:fC-/ in Philadelphia English*

input	optimal output	IDENT-length	*ǣf _σ]	*ǣ
/æfgən/	[æf.gən]		*	
/æ:ftəɪ/	[æ:f.təɪ]			*

By the same token, the following rankings capture the fact that /f/ is a trigger of lexically gradual tensing in both New York City and Philadelphia, whereas /g/ is a trigger only in New York City; see (11).

⁹ Labov (1989, 1994) provides no information on the Philadelphian pronunciation of the second syllable of *Afghan*. Donka Minkova (personal communication) informs me that the second vowel is rarely reduced in American English. This point is not essential to the discussion below.

¹⁰ I do not claim theoretical validity for these markedness constraints, but use them here merely for the purposes of illustration. Further phonetic, phonological, and typological investigation would be needed to ascertain the precise nature of the constraints involved. In the labels, I replace the IPA length symbol (:) with the classical macron (ˉ) and breve (˘) just to remind the reader that *ǣf_σ] specifically bans *short* [æ] before a tautosyllabic [f].

- (14) (a) Philadelphia IDENT-length » *ǣf_σ] » *ǣ » *ǣg_σ]
 (b) New York City IDENT-length » *ǣf_σ], *ǣg_σ] » *ǣ

I suggest that a learner who has acquired the constraint hierarchy in (14a) will be biased to expect input /æ:/ in the environment /__fC/, but input /æ/ in the environment /__gC/. Given this bias, any phonetic effect impinging on the realization of the contrast between /æ/ and /æ:/ may cause learners to fail to acquire input /-æfC-/ for particular lexical items, substituting unmarked /-æ:fC-/ instead. Similarly, a learner who has mistakenly acquired a lexical representation with /-æ:fC-/ for a lexical item *x* will tend not to recover from this error, except upon massive exposure to tokens of *x* with /æ/. Labov (1994:518-526) provides empirical evidence that the lexical incidence of /æ/ and /æ:/ in Philadelphia English is indeed difficult to acquire: children born in Philadelphia to out-of-state parents typically fail to learn the native Philadelphian distribution, presumably because they initially acquire non-Philadelphian lexical entries from their parents and then fail to recover from the error.

As it stands, this account of the top-down factors driving classical lexical diffusion remains incomplete (Goldsmith 1995a). The basic idea is that markedness constraints, even when crucially dominated and therefore unable to cause unfaithfulness, nonetheless exert an indirect pressure on learners to switch features to their unmarked value in input representations. As Goldsmith points out, however, this incorrectly predicts that all lexical contrasts should be vulnerable to loss by diffusion, insofar as every feature has an unmarked value in every context. Goldsmith suggests that, in fact, the pressure towards category substitution in lexical representations is only felt in cases of *marginal contrast*. In the case of Philadelphia *æ*-tensing, Labov's (1989:45) data strongly corroborate this insight. Labov defines two word classes:

- (i) The 'normally tense' class consists of words containing environments where *æ* is nearly predictably tense. Only 2.9% of word tokens in this class show /æ/.
- (ii) The 'normally lax' class consists of words containing environments where *æ* is nearly predictably lax. Only 1.2% of word tokens in this class show /æ:/.

Strikingly, the 'normally tense' and 'normally lax' classes together account for 91.8% of Labov's data: only 443 out of 5,373 word tokens belong to the residual category, where the tenseness or laxness of *æ* cannot be predicted with more than 97% accuracy. In this light, it seems desirable to build Goldsmith's notion of marginal contrast into our account of classical lexical diffusion: one should probably say that learners are biased to replace a marked category with an unmarked one in lexical representations when the two categories are only marginally contrastive. Further research should determine the basis of this effect by ascertaining how marginal contrastivity impacts on lexical representation.

Pace Phillips (1998), this approach to the causes of classical lexical diffusion enables one to make correct predictions regarding the impact of token frequency on the progress of diffusing changes. These predictions depend crucially on the recognition that classical lexical diffusion is actuated by a combination of top-down and bottom-up factors:

- (i) In the case of the Mid-Atlantic dialects of American English, for example, I have suggested that phonetic factors tending to compromise the realization of the contrast between /æ/ and /æ:/ will cause learners to fail to acquire /æ/ in /æ:/-favouring environments. We know, however, that gradient phonetic effects such as coarticulation and gesture reduction are more pronounced in items with high token frequency.¹¹ From this it follows that the probability of substitution of /æ:/ for /æ/, insofar as it depends on phonetic effects, will be greater for high-frequency words.
- (ii) On the other hand, I have suggested that, because of the bias introduced by the constraint hierarchy (in circumstances of marginal contrastivity), children will tend not to recover from overgeneralizations in the distribution of /æ:/, unless assisted by massive exposure to /æ/ in tokens of the relevant words. This predicts that the words with the very highest token frequency may exceptionally withstand the change.

These predictions exactly match Labov's findings about Philadelphia *æ*-tensing. He reports that, 'In the two cases of change in progress, more frequent words are more frequently selected. [...] Yet frequency exhibits only a general correlation with tensing and fails to account for the fact that the most common words [...] show the least tendency to shift to the tense class' (Labov 1989:44).

21.4 The view from functionalist phonology

21.4.1 Exemplar clouds

As I pointed out in the conclusion of §21.2, there has lately been a vigorous reaction against the classical modular approach to the phonology-phonetics interface (2) and, with it, against the received view of the implementation of phonological change (1). There is of course a wide range of opinion among the dissenters, but one can nonetheless identify a coherent research paradigm coalescing around a few programmatic drives: usage-based functionalism, phonetic reductionism, and connectionism. Bybee (2001) provides a synthetic statement of this research programme, which for convenience I shall call 'functionalism' *tout court*.

Essential to current functionalist thinking is the idea that the lexicon consists of a vast repository of highly-detailed memory traces of phonetic episodes experienced by the speaker: so-called 'exemplars' (Johnson 1997). Exemplars are linked to one another by a network of connections based on similarity in a high-dimensional phonetic space. Crucially, phonological categories do not exist independently of the exemplars. This idea comes in several versions. In the strongest version, categories exist implicitly in the patterns of connection between exemplars: a category, in this sense, is no more than a cloud of similar exemplars connected in some dimension. In a more conciliatory reading, categories exist explicitly as labels attached to the exemplars, and could even be accessed by rules referring to typed variables (see §21.4.3).¹² In either case, association with

¹¹ The reasons for this phenomenon, and its implications for diachronic phonology and for the phonology-phonetics interface, are discussed in §21.4 below.

¹² On the need for typed variables in linguistic rules, see Marcus (2001) and Jackendoff (2002).

exemplar clouds endows categories with prototype structure: tokens of the category may be more or less central, and categorial boundaries are fuzzy. Finally, lexical representation is continuously updated in performance, as new exemplars are stored in long-term memory and old exemplars decay.

From this position, Bybee (2001:40-41) suggests that many – if not all – sound changes are simultaneously *gradient and diffusing*. This assertion roundly contradicts the predictions of the classical modular architecture, where gradient diffusion is precisely the only mode of implementation ruled out in principle: see (4).

Conceptually, Bybee's assertion rests on the assumption that every lexical item is associated with its own exemplar cloud. The phonetic properties of each lexical item shift as new exemplars are added to its cloud during language use. Accordingly, sound change must be phonetically gradual, insofar as it involves a continuous shift in the aggregated phonetic properties of the cloud. It must also be lexically gradual, since each lexical item has its own pattern of use, recorded in its own exemplar cloud.

Empirically, Bybee draws support from the observation that lexical items with high token frequency display greater amounts of coarticulation and gestural overlap than low frequency items. In a well-known example, the average duration of the medial [ə] in a high-frequency word such as *nursery* ['nɜːsəri] turns out to be shorter than in a low-frequency word such as *cursor* ['kɜːsəri] (Bybee 1998:68).

(15) *Effect of token frequency on gradient phonetic patterns*

<i>Token frequency</i>	<i>Amount of coarticulation and reduction</i>	<i>Example</i>
high	high	shorter [ə] in <i>nursery</i>
low	low	longer [ə] in <i>cursor</i>

Bybee infers that, as predicted by the exemplar model, the lexical representations of *nursery* and *cursor* must contain detailed quantitative information about degrees of gestural reduction and overlap.

21.4.2 *The problem of phonetic residue*

According to Bybee, all changes are lexically gradual, including those actuated by phonetic factors, whether articulatory or perceptual. However, diffusing changes often become arrested before completion, leaving behind a residue of unaltered words (Wang 1969). Accordingly, Bybee's claim leads to some surprising predictions. For example, endogenous lexical splits should be commonplace: an instance of a lexical split is the unpredictable evolution of Middle English short /a/ in the Mid-Atlantic region of the United States (see §21.2 and §21.3.3); this split would be described as 'endogenous' if it had not been actuated by contact (cf. below). Moreover, gradient diffusion predicts that, over time, the lexicon will preserve remnants of old phonemes and exceptions to new allophonic patterns, all left behind by arrested changes. Indeed, if holistic phonetic targets were kept in long-term memory in quite the same way as lexicalized

morphological constructs, then phonetic relics should be as unremarkable as stored morphological irregularities like *children*, *oxen*, *feet*, or *wolves*.¹³

In fact, lexical splits are rare and typically arise in contact situations. For example, contact appears to have played an important rôle in triggering the lexical split of short /a/ in southern British English during the Early Modern period (Labov 1989:§2). At this point in the history of the language, native Middle English /a:/ had already started on its way to /eɪ/ by the Great Vowel Shift. However, native short /a/ developed a lengthened allophone before coda /f, θ, s/; cf. (11). When this new allophonic [a:] merged with the long vowel present in some French loans such as *France* and *dance*, its distribution became marginally contrastive, and so the conditions arose for the diffusing spread of a new /a:/ phoneme. The Mid-Atlantic version of æ-tensing discussed in §21.2 and §21.3.3 ultimately descends from this southern British split (see Labov 1994:529ff.). Similarly, *pace* Chen (1972), Wang and Lien (1993) concede that the remarkable lexical split of Middle Chinese tone III in Chao-zhou (§21.2) was actuated by prolonged contact between an implanted literary dialect and an indigenous colloquial dialect; see further Labov (1994:451-454).

In response to these problems, Bybee (1998:72; 2001:54) cursorily suggests that speakers are driven by efficiency requirements to constantly reuse a finite set of highly practised neuromotor programmes in speech. The pervasive redeployment of these articulatory plans prevents left-over junk from accumulating in phonological systems. This proposal is eminently reasonable: indeed, the modular feedforward architecture diagrammed in (2) is one of the possible instantiations this idea. However, Bybee does not fill in the detail. What sizes do these reusable units come in? How exactly do they relate to the episodic memory traces in the lexicon? Do they exist, for example, as labelled pieces of the exemplars themselves? If so, how exhaustively is each exemplar labelled at each level of granularity? Given certain plausible answers to these questions, Bybee's proposal becomes very similar to the hybrid grammatical model proposed by Pierrehumbert (2002), in which each lexical item is associated both with a categorical phonological parse and with a phonetic exemplar cloud (see §21.4.3). This type of hybrid model avoids the problem of phonetic residue much more effectively.

21.4.3 *Dealing with frequency effects*

One of Bybee's key empirical arguments in favour of gradient diffusion is the sensitivity of coarticulation and gestural reduction to the token frequency of lexical items: see (15). How serious a problem is this for modular feedforward models of phonology? Significantly, both Pierrehumbert (2002:§3) and Coleman (2003:§5.4), though sympathetic to Bybee's position, concede that the classical modular architecture can be readily modified to accommodate such effects. Functionally, the correlation shown in (15) makes perfect sense: in order to facilitate the task of lexical recognition for the listener, speakers shift towards hyperspeech, with less coarticulation and reduction, when uttering words that are hard to access. Low-frequency words, with their low resting activation, fall into this group, as do words with low contextual predictability (Jurafsky *et*

¹³ See Kiparsky (1988:366) for discussion of Bloomfield's (1933) formulation of this argument.

al. 2001) and words with high neighbourhood density (Wright 2003).¹⁴ Models like (2) could therefore deal with frequency effects by enriching phonological representations with information about lexical accessibility, which would thus be made available to the phonetic module. For example, Pierrehumbert (2002:107) suggests that, when a morph is inserted into a phonological expression, the prosodic-word node that hosts the morph could be annotated with a numerical index of lexical accessibility; in the phonetic module, coarticulation and reduction rules would lower the expenditure of articulatory effort on elements with a high accessibility index. Under this proposal, the lexicon remains free of phonetic detail.¹⁵

Nonetheless, Pierrehumbert (2002) advocates a more evenly balanced compromise between modular and exemplar-based models (see Harris [6.2.1]). In line with traditional approaches, she proposes that, for each linguistic expression, a phonological processor operating symbolic rules constructs a phonological representation consisting of discrete categories. Each of these discrete phonological categories, however, is associated with an exemplar cloud: in production, a phonological category is assigned a phonetic realization target by making a random selection from its cloud. In an utterance of the word *nursery*, for example, a duration target for the medial /ə/ is set by randomly selecting exemplars of /ə/ and calculating their average duration. Crucially, the relative contribution of each selected exemplar to the production target is weighted: e.g. in the production of *nursery*, exemplars of /ə/ located in memory traces of the word *nursery* count for more than exemplars of /ə/ located in memory traces of other words (such as *cursor*). In this proposal, therefore, the highly detailed word-specific phonetic information contained in episodic memory does not supply holistic production targets for lexical items, but rather introduces subtle biases in the phonetic implementation of discrete phonological representations. Thus, “word-specific phonetic effects are second-order effects” (Pierrehumbert 2002:134); that is, “the influence of particular words on phonetic outcomes is secondary, with the actual phonological makeup of the words providing the primary influence” (Pierrehumbert 2002:129).

This integrative approach strikes me as worth pursuing. On a general note, Pierrehumbert’s (2002) proposals may find a natural home in a cognitive paradigm like Karmiloff-Smith’s (1994) modified constructivism: in Pierrehumbert’s model, the acquisition of phonology must involve, *inter alia*, a process of categorical labelling of episodic memory traces, which can easily be conceptualized as representational redescription in Karmiloff-Smith’s sense (see §21.3.2 above). Also relevant here is Pinker’s (1999:279) assertion that rules and exemplars coexist in human cognition as ‘two ways of knowing’. Pursuing these links, one realizes that diachronic phonology provides us with a unique window on the nature of the mind.

¹⁴ The phonological neighbourhood density of a word *x* is said to be high if the lexicon contains many words that are phonologically similar to *x*.

¹⁵ The lexical accessibility index would be an attribute with gradient values (cf. (3a) above), but, crucially, it would not directly encode phonetic properties.

References

- Allen, William Sydney (1973). *Accent and rhythm*. Cambridge: Cambridge University Press.
- Anderson, Stephen R. (1988). Morphological change. In Newmeyer (1988). 324-362.
- Archangeli, Diana (1984). *Underspecification in Yawelmani phonology and morphology*. Doctoral dissertation, MIT. Published (1988) New York: Garland Press.
- Baudouin de Courtenay, Jan (1895). An attempt at a theory of phonetic alternations. [Translation of *Versuch einer Theorie phonetischer Alternationen: ein Kapitel aus der Psychophonetik*.] In Edward Stankiewicz (ed.) (1972). *A Baudouin de Courtenay anthology*. Bloomington: Indiana University Press. 144-212.
- Bermúdez-Otero, Ricardo (1998). Prosodic optimization: the Middle English length adjustment. *English Language and Linguistics* 2. 169-197.
- Bermúdez-Otero, Ricardo (forthcoming a). Phonological change in Optimality Theory. In Keith Brown (ed.) *Encyclopedia of language and linguistics*, 2nd edition. Oxford: Elsevier.
- Bermúdez-Otero, Ricardo (forthcoming b). *Stratal Optimality Theory*. Oxford: Oxford University Press.
- Bermúdez-Otero, Ricardo & Kersti Börjars (2006). Markedness in phonology and in syntax: the problem of grounding. In Patrick Honeybone & Ricardo Bermúdez-Otero (eds.) *Linguistic knowledge: perspectives from phonology and from syntax*. Special issue, *Lingua* 116:2.
- Bermúdez-Otero, Ricardo & Richard M. Hogg (2003). The actuation problem in Optimality Theory: phonologization, rule inversion, and rule loss. In D. Eric Holt (ed.) *Optimality Theory and language change*. Dordrecht: Kluwer. 91-119.
- Bermúdez-Otero, Ricardo & April McMahon (forthcoming). English phonology and morphology. In Bas Aarts & April McMahon (eds.) *The handbook of English linguistics*. Oxford: Blackwell.
- Bloomfield, Leonard (1933). *Language*. New York: Holt.
- Burton-Roberts, Noel, Philip Carr & Gerard Docherty (eds.) (2000). *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press.
- Bybee, Joan (1998). The phonology of the lexicon: evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.) *Usage-based models of language*. Stanford: CSLI Publications. 65-85.
- Bybee, Joan (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Chen, Matthew (1972). The time dimension: contribution toward a theory of sound change. *Foundations of Language* 8. 457-498.
- Cheng, Chin-Chuan & William S-Y. Wang (1973). Tone change in Chao-zhou Chinese: a study in lexical diffusion. In Braj B. Kachru, Robert B. Lees, Yakov Malkiel, Angelina Pietrangeli & Sol Saporta (eds.) *Issues in linguistics: papers in honor of Henry and Renée Kahane*. Urbana: University of Illinois Press. 99-113.
- Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

- Coleman, John (2003). Commentary: probability, detail and experience. In Local *et al.* (2003). 88-100.
- Croft, William (2000). *Explaining language change: an evolutionary approach*. Harlow: Pearson Education.
- Dressler, Wolfgang (1985). *Morphology: the dynamics of derivation*. Ann Arbor: Karoma.
- Fodor, Jerry A., & commentators (1985). Précis and open peer commentary of *The modularity of mind*. *Behavioral and Brain Sciences* **8**. 1-42.
- Fox, Anthony (1995). *Linguistic reconstruction: an introduction to theory and method*. Oxford: Oxford University Press.
- Fromkin, Victoria A. (1971). The nonanomalous nature of anomalous utterances. *Lg* **47**. 27-52.
- Goldsmith, John A. (1995a). Phonological theory. In Goldsmith (1995b). 1-23.
- Goldsmith, John A. (ed.) (1995b). *The handbook of phonological theory*. Oxford: Blackwell.
- Gordon, Matthew J. (2001). *Small-town values and big-city vowels: a study of the Northern Cities Shift in Michigan*. (Publications of the American Dialect Society **84**.) Durham, NC: Duke University Press.
- Hayes, Bruce (2000). Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: phonology, syntax, and acquisition*. Oxford: Oxford University Press. 88-120.
- Hayes, Bruce & Steriade, Donca (2004). Introduction: the phonetic bases of phonological markedness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*. Cambridge: Cambridge University Press. 1-33.
- Hyman, Larry M. (1976). Phonologization. In Alphonse Juilland, with A. M. Devine & Laurence D. Stephens (eds.) *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*. Saratoga: Anma Libri. 407-418.
- Itô, Junko & R. Armin Mester (1995). Japanese phonology. In Goldsmith (1995b). 817-838.
- Jackendoff, Ray (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Janda, Richard (1999). Accounts of phonemic split have been greatly exaggerated—but not enough. *Proceedings of the International Congress of Phonetic Sciences* **14**. 329-332.
- Johnson, Keith (1997). Speech perception without speaker normalization. In Keith Johnson & John W. Mullennix (eds.) *Talker variability in speech processing*. San Diego: Academic Press. 145-165.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. 229-254.
- Karmiloff-Smith, Annette, & commentators (1994). Précis and open peer commentary of *Beyond modularity: a developmental perspective on cognitive science*. *Behavioral and Brain Sciences* **17**. 693-745.

- Keating, Patricia A. (1988). The phonology-phonetics interface. In Newmeyer (1988). 281-302.
- Kiparsky, Paul (1971). Historical linguistics. In W. O. Dingwall (ed.) *A survey of linguistic science*. College Park: University of Maryland Linguistics Program. 576-642. Reprinted in Paul Kiparsky (1982). *Explanation in phonology*. Dordrecht: Foris. 57-80.
- Kiparsky, Paul (1982). Lexical morphology and phonology. In I.-S. Yang (ed.) *Linguistics in the morning calm: selected papers from SICOL-1981*. Seoul: Hanshin. 3-91.
- Kiparsky, Paul (1988). Phonological change. In Newmeyer (1988). 363-415.
- Kiparsky, Paul (1995). The phonological basis of sound change. In Goldsmith (1995b). 640-670.
- Labov, William (1981). Resolving the Neogrammarian controversy. *Lg* 57. 267-308.
- Labov, William (1989). Exact description of the speech community: short *a* in Philadelphia. In Ralph W. Fasold & Deborah Schiffrin (eds.) *Language change and variation*. Amsterdam: John Benjamins. 1-57.
- Labov, William (1994). *Principles of linguistic change: internal factors*. Oxford: Blackwell.
- Labov, William (2001). *Principles of linguistic change: social factors*. Oxford: Blackwell.
- Ladd, D. Robert & James M. Scobbie (2003) External sandhi as gestural overlap? Counter-evidence from Sardinian. In Local *et al.* (2003). 164-182.
- Lindblom, Björn (1990). Explaining phonetic variation: a sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer. 403-439.
- Local, John, Richard Ogden & Rosalind Temple (eds.) (2003). *Phonetic interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
- Marcus, Gary (2001). *The algebraic mind*. Cambridge, Mass.: MIT Press.
- Martinet, André (1960). *Eléments de linguistique générale*. Paris: Colin. English translation by Elisabeth Palmer (trans.) (1960). *Elements of general linguistics*. Chicago: University of Chicago Press.
- McCarthy, John J. (2002). *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
- McMahon, April (2000). *Lexical phonology and the history of English*. Cambridge: Cambridge University Press.
- Myers, Scott (2000). Boundary disputes: the distinction between phonetic and phonological sound patterns. In Burton-Roberts *et al.* (2000). 245-272.
- Newmeyer, Frederick J. (ed.) (1988). *Linguistics: the Cambridge survey*. Volume 1: *Linguistic theory: foundations*. Cambridge: Cambridge University Press.
- Ohala, John J. (1989). Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.) *Language change: contributions to the study of its causes*. Berlin: Mouton de Gruyter. 173-198.
- Osthoff, Hermann & Karl Brugmann (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Volume 1. Leipzig: S. Hirzel. English translation of the preface in Winfred Lehmann (ed., trans.) (1967). *A*

- reader in nineteenth century historical Indo-European linguistics*. Bloomington: Indiana University Press. 197-209.
- Phillips, Betty S. (1998). Lexical diffusion is *not* lexical analogy. *Word* **49**. 369-380.
- Pinker, Steven (1999). *Words and rules: the ingredients of language*. London: Weidenfeld & Nicolson.
- Pierrehumbert, Janet (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology 7*. Berlin: Mouton de Gruyter. 101-139.
- Pierrehumbert, Janet, Mary E. Beckman & D. R. Ladd (2000). Conceptual foundations of phonology as a laboratory science. In Burton-Roberts *et al.* (2000). 273-303.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published with revisions (2004). Oxford: Blackwell.
- Sproat, Richard & Osamu Fujimura (1993). Allophonic variation of English /l/ and its implications for phonetic implementation. *JPh* **21**. 291-311.
- Steriade, Donca (1995). Underspecification and markedness. In Goldsmith (1995b). 114-174.
- Wang, William S-Y. (1969). Competing changes as a cause of residue. *Lg* **45**. 9-25.
- Wright, Richard (2003). Factors of lexical competition in vowel articulation. In Local *et al.* (2003). 75-87.
- Zsiga, Elizabeth C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Bruce Connell & Amalia Arvaniti (eds.) *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press. 282-302.