

Phonologization*

Paul Kiparsky
Stanford University

*Some of this material was presented at the San Francisco LSA meeting in January 2002.

1 Phonologization in structuralist phonology

If there is a phonemic level of representation, the question arises how and why phonemes originate and are lost. This is the problem of PHONOLOGIZATION and MERGER. The structuralists held that allophones are phonologized, i.e. become phonemic, when their conditioning environment of is eliminated by sound change. The idea was introduced by V. Kiparsky (1932) and Twaddell (1938),¹ who both illustrated it with the phonologization of the front rounded umlaut vowels *ö* and *ü* in German. They argued that the umlaut vowels became distinctive when the *i* or *j* that conditioned them was reduced or deleted, at which point lexical representations were restructured with the former allophonic variants as phonemes.²

- (1) a. Old High German (OHG):
- Nom.Pl. /huot-i/ → *hüeti* ‘hats, helmets’
 - Dat.Sg. /huot-e/ → *huote*
- b. Middle High German:
- Sound change: *hüeti* > *hüete*
 - Restructuring: /huot-i/ > /hüet-e/

This mechanism, called SECONDARY SPLIT (Hoenigswald 1965, see Cser and Fox, this volume), has been standardly assumed to account for phonologization. However, it leaves two questions unanswered. First, when the conditioning environment goes away (as here by reduction of the full vowels to *-e* or to *-ə*), *why* do its effects remain?³ Secondly, why does the loss of a conditioning environment not *always* cause phonologization? Why do the conditioned allophones sometimes just go away?⁴ Let’s call them the PHONOLOGIZATION PROBLEM and the NON-PHONOLOGIZATION PROBLEM, respectively.

Both problems are deep ones, and reach the foundations of phonological theory. The phonologization problem has been at the core of phonological theorizing, but has no generally agreed on solution. The non-phonologization problem has barely begun to be discussed. I begin by reviewing the two main proposed solutions to the phonologization problem, and suggest a new one based on Lexical Phonology and more particularly its OT version (Stratal OT), which I argue also resolves the non-phonologization problem.

The best-known proposal for resolving the phonologization problem is due to Saussure and Bloomfield. It depends on a paradoxical marriage of synchronic structuralism to diachronic neogrammarianism. The key idea is that a sound change is located outside of the linguistic system that it

¹Kiparsky’s article is an introduction to synchronic and historical structural phonology, written after his stay in Prague where the new theory was just then taking shape. It correctly avoids Twaddell’s assumption that OHG orthography was phonemic. Kiparsky’s well-meaning academic mentors eventually managed to dissuade the young graduate student from continuing his forays into phonological theory and other such modish nonsense.

²The Old High German data cited in this article can be found in Braune & Reiffenstein 2004 or in any standard grammar of the language. In accord with orthography and standard practice in Germanic linguistics I write the front rounded vowels as *ü*, *ö*.

³“Why did the front vowels not become back again, why did the frontness stay, once the influence of /i j/ was removed?” (Lieberman 1991: 126).

⁴“Why do allophones sometimes remain and other times revert?” (King 1971: 4).

will transform. A less-known but increasingly popular alternative is to enrich the phonology with phonetic information, abandoning the concept of a phoneme as a contrastive entity, and positing that phonemes-to-be somehow get phonologized *before* they become contrastive through the loss of the conditioning factor.

After reviewing these solutions in turn, I will argue for a version which combines aspects of both, implemented in the Stratal OT framework.

For Saussure, the basic fact about language is the ARBITRARINESS OF THE SIGN. Among its consequences are a sharp divide between synchrony and diachrony. This division is *not* merely a methodological one, still less a practical one or one based on conventional boundaries between academic subdisciplines. It is a *conceptually necessary* consequence of arbitrariness.

Linguistics must be separated in two. There is an irreconcilable duality, created by the very nature of things [...] in systems of values. (Saussure 1993:104, cf. Saussure 1914/1959:79-80)

A historical event such as a sound change *qua* phonetic mutation, and the restructuring of the phonological system it may cause, are totally different things: a sound change can bring about radical discontinuities, or it could have no effect on the system whatever.

There is no inner bond between the initial fact [the phonetic change] and the effect that it may subsequently produce on the whole system [phonology or grammar]. (Saussure 1914/1959:87)

The sharp segregation of historical change from synchronic structure — call it SAUSSURE'S FIREWALL — implies that, although everything in grammar is interrelated as a system, *sound change has no access to that system*. Blindly and structure-independently, it alters merely the material implementation of speech. The abstract synchronic system, characterized by networks of relations and systems of constraints, is affected only indirectly by those alterations. For Saussure, the synchronic constraints in the mind of the speaker and the historical processes that modify the articulation of speech are not only ontologically, but also formally distinct. Constraints are GENERAL (transparent, or inviolable, in current terminology), whereas processes are ACCIDENTAL and PARTICULAR. Constraints are PRECARIOUS (they could be overturned by the next change), while processes are IMPERATIVE (sound change is exceptionless).

Back to the example of umlaut: the phonetic mutations of umlaut and reduction/syncope both altered the physical aspect of speech, but they had very different effects on the system. The structural reflex of umlaut was purely *phonetic*: it introduced the constraint “no back vowels before *i*”. This is an allophonic distribution and has no bearing on the phonological system. Vowel reduction and subsequent syncope had no phonetic repercussions on the umlaut vowels, but an all the more drastic impact on their *phonological* status: it caused them to be reanalyzed as distinct phonemes. The new phonetic givens lead to a restructured phonological system with new phonemes /ü/, /ö/ and a new constraint “no unstressed full vowels”. The site of phonemic contrast has shifted one syllable to the left.

This is a consistent theory of sound change. But the dualist ontology of Saussure's Firewall is a heavy price to pay. It excludes all structural explanations for sound changes and for constraints

on sound change of the sort pioneered by Jakobson and since pursued in different ways by Martinet, Labov and others, and, still differently, in generative and OT work. In particular, it makes inexplicable the fact that sound change never subverts phonological universals.

Another objection to this theory is that it offers no solution to the non-phonologization problem. Why does the predicted secondary split sometimes not happen, and the conditioned allophones just disappear? For example, vowel fronting in various dialects of English (e.g. in *calf*, *goat*, *cough*), and vowel backing (*girl*, dialectally in *kit*) usually don't produce contrasts between front and back *k* such as structuralist doctrine predicts should arise (an exception is the Jamaican English contrast *cat* [kʲat] vs. *cot* [kat] or [kɔt], Wells 1982: 569).⁵ Such cases were discussed under the heading of "rule insertion" in generative theorizing on sound change (King 1973).

Another type of rule insertion, also problematic for the structuralist account of secondary split⁶ is that a sound change can interact with, and be constrained by, existing phonological processes and constraints in the language. King notes that Old English syncope of unstressed *e* in words like *bindest* 'you bind' and *bidest* 'you ask' feeds the previously existing voicing assimilation rule, so that the outcome is *bintst*, *bitst*. Saussure's Firewall here predicts that syncope should extend the voicing opposition to what was until then a neutralizing assimilation environment, creating a contrast between previously existing assimilated clusters such as *-ts-* and new clusters from syncope such as **-ds-*. This is certainly a conceivable outcome, but it is not what happened in Old English; there is no reason to believe that clusters such as **-ds-* ever existed, even immediately after syncope.

The third problem is that sound changes can be *blocked* by existing synchronic constraints. For example, syncope can fail to apply just in those cases where it would create a prohibited stress configuration (e.g. a lapse or clash), or a prohibited syllable structure or foot structure. In English, the variable pre-sonorant syncope in words like *generative* is inapplicable before a stressed syllable, as in *generate* (**gen'rate*), where it would produce back-to-back stresses, which are disfavored in English. Technically, such conditions on sound changes can be specified as conditioning factors, but only at the cost of a loss of the generalization that the conditioning factors are manifestations of active phonological constraints of the language.

The first two types of problematic cases are the historical analogs of the two types of transparent rule interaction in synchronic phonology: vowel backing in *kit* BLEEDS *k*-fronting, and syncope in *bidest* FEEDS voicing assimilation. The third type of problematic case also involves transparent interaction, in the sense that sound change avoids creating surface exceptions to a constraint that is operative in the language.

In short, sound changes can interact transparently with existing processes. Such transparent interactions can involve feeding or bleeding by the sound change, or blocking of the sound change by a constraint. Alongside such transparent interactions, sound change can also result in opacity, which in terms of change means phonologization and the creation of new contrasts. Structuralist

⁵Commenting on a proposal that umlaut vowels disappeared in Scandinavian when the triggering front vowels were syncopated, Benediktsson (1982: 9) states: "The principle that phonetic variants, in consequence of the conditioning factors, may 'revert to the neutral starting-point' [...], though perhaps consistent with generative theory, seems hardly compatible with those of structural phonology; at any rate, if it is accepted, the principle of phonemicization is then reduced to an ad-hoc postulate, of little or no explanatory value."

⁶As well as for the theory espoused by Blevins 2004, see Kiparsky 2006.

historical phonology has privileged the latter scenario to the point of all but ignoring the well-documented possibility of transparent interaction.

Post-structuralist theories which relate historical and synchronic phonology have been unable either to replicate Saussure's Firewall without some extrinsic stipulation, or to derive the generalization that it is intended to capture in some other way. Classical generative grammar's unification was accomplished by generalizing the processual approach and modeling sound change as the addition of rules. The question then arose where they are added. Obviously rules cannot be added anywhere, but saying that they are added to the end of the grammar is arbitrary, and in any case fails to do justice to the cases of "rule insertion".

King (1971) argued that sound changes interact transparently only with "phonetic rules" — the "trivial case" of rule insertion, as he called it. His observation has held up well; the "non-trivial" cases have been fairly convincingly explained away (see Jasanoff 2003). It is fair to assume that the "phonetic rules" of King's generalization are a language's postlexical rules. If so, we can rephrase his generalization like this:

- (2) a. SECONDARY SPLIT: Sound changes render lexical processes opaque.
- b. BLOCKING AND "RULE INSERTION": Sound changes interact transparently with postlexical processes.

The second, less well explored approach to the phonologization problem assumes that prospective phonemes are already phonologized by the time they become contrastive (Ebeling 1960, Korhonen 1969, Liberman 1991). For example, if the umlaut vowels are already phonemes (or QUASI-PHONEMES, as Korhonen calls them) before the *-i-* that conditions them is lost, then they would naturally remain unaffected by the latter sound change.

There is evidence that non-contrastive allophones can be internalized in the lexical phonology (e.g. Hellberg 1978, 1980), but a theory of phonologization must specify when and why this happens. We must be able to tell that an allophone has become a quasi-phoneme independently of the post hoc information that it is phonologized when another sound change occurs. It has been suggested that features tend to be phonologized if they belong to a feature class which is already distinctive (Kiparsky 1988). Though generally consonant with observations about priming effects in sound change, this idea is not precise enough to make predictions about when phonologization will take place. Another suggestion, made by Janda (2003: 413) in a vigorous plea for early phonologization, is that allophones become quasi-phonemes "for reasons having to do with phonetic distance", though he does not say how much distance, and on what dimension, or cite evidence that distance matters at all.

A starting point for a more substantive theory of phonologization might be Jakobson's observation that allophonic properties can become perceptually more salient than the phonemic ones that condition them (Jakobson, Fant, and Halle 1952). Russian [i̯] and [i] are allophones of /i/ after respectively back and front consonants, yet the allophonic vowel distinction is a more salient cue to the contrast than the phonemic consonantal one (especially in the case of sibilants because of their high-frequency noise, e.g. /si/ [si̯] and /s'i/ [s'i]). Related to this perceptual saliency of the vowels, as Jakobson pointed out, is the fact that [i̯] and [i] are perceived as categorically distinct elements, and even reified in the metalinguistic terms [ikat̚] 'to pronounce [i̯]' and [ikat] 'to pronounce [i]'.

The vowels [i̯] and [i] are like two phonemes in that any unrounded high vowel token is assigned to one or the other type; perceptually they are two distinct categories. Other Russian vowels are also strongly affected by palatalization: e.g. /a/ is fronted towards [æ] to varying degree before, after, and most of all between palatalized consonants, but the allophones are apparently not categorically perceived as belonging to two types; correspondingly there is no *[ækatʲ] ‘to pronounce [æ]’.

A plausible hypothesis is that allophones become quasi-phonemes when they become governed by categorical rather than gradient constraints (Flemming 2001), and acquire greater perceptual salience than their conditioning environments. How are these two properties related to each other, and how we can build a theory of phonologization on them?

We cannot build on structuralist phonology here because it attributes categoriality and saliency to phonemic representations. It treats feature specifications at the phonemic level as categorical, though allowing allophonic/postlexical feature specifications to be gradient. On this view, phonemic representations specify exactly the invariant distinctive features of the language. Quasi-phonemes are not allowed at the phonemic level since redundant, predictable feature values are excluded from lexical representations. An approach that does provide a theoretical basis for the quasi-phoneme is Stratal OT.

2 Parallel OT

Constraint-based theories such as OT eliminate processes in favor of constraints, and model sound change as the promotion of markedness constraints. For present purposes it is important to distinguish parallel and stratal versions of OT (see also Holt, this volume). Stratal OT distinguishes levels (strata), and claims that constraints operate transparently within a level, but that the levels themselves interact much as ordered rules did in SPE phonology. The generalizations in (2) are then predicted. Promotion of a postlexical constraint will lead to non-phonologization (blocking and rule insertion) effects. Blocking arises when the promoted postlexical constraint is dominated within the postlexical phonology by an antagonistic constraint (e.g. syncope by a restriction on syllable structure). “Rule insertion” (a misnomer in this framework, of course) arises when the promoted postlexical constraint winnows away candidates that would otherwise emerge as winners by the lower-ranking postlexical constraints.

Phonologization (secondary split), on the other hand, takes place because constraints at a given level do not affect the operation of constraints at earlier levels. It arises when the context of a lexical process is made opaque by an innovated postlexical process, that is, by a sound change qua promoted markedness constraint. From the perspective of Stratal OT, then, the reason why the umlaut vowels became phonologized when the triggering context was lost is that they were introduced in the lexical phonology.

Before going into the details, let us emphasize that this solution is not available in parallel OT. As far as I can see, parallel OT actually has no coherent characterization of secondary split, for reasons which are homologous to its failure to deal with opacity. To see why, consider a bare-bones OT constraint system for the pre- and post-phonologization stage of umlaut.

- (3) a. IDENT(Hi): the input and output values of [High] are identical.
- b. **ü*, **ö*: rounded vowels are back.

- c. REDUCE: no full (unreduced) unstressed vowels.
- d. IDENT(Front): the input and output values of [Front] are identical.
- e. AGREE(Front): no back vowels before *i, j* (the constraint that enforces umlaut).

The original grammar has the same ranking in all phonological strata; crucially IDENT(Hi) » REDUCE, and *ü,*ö » IDENT(Front), AGREE(Front).

(4)

Input	Stage 0: No umlaut					
	Output	IDENT(Hi)	*ü, *ö	REDUCE	IDENT(Front)	AGREE(Front)
uCi	☞ uCi			*		*
	üCi		*	*	*	
	uCe	*				
	üCe	*	*		*	
uCe	uCi	*		*		*
	üCi	*	*	*	*	
	☞ uCe					
	üCe		*		*	
üCe	uCi	*		*	*	*
	üCi	*	*	*		
	☞ uCe				*	
	üCe		*			

The first sound change introduces allophonic umlaut into the language:

(5)

Input	Stage 1: allophonic umlaut					
	Output	AGREE(Front)	IDENT(Hi)	*ü, *ö	REDUCE	IDENT(Front)
uCi	uCi	*			*	
	☞ üCi			*	*	*
	uCe		*			
	üCe		*	*		*
uCe	uCi	*	*		*	
	üCi		*	*	*	*
	☞ uCe					
	üCe			*		*
üCe	uCi	*	*		*	*
	üCi		*	*	*	
	☞ uCe					*
	üCe			*		

Now vowel reduction takes place: an innovating grammar in which REDUCE outranks IDENT(Hi) begins to compete with (5) and eventually supplants it. But the new output *üCe* cannot be accounted

for. No matter how the remaining constraints are ranked, the reduction causes umlaut to be undone. The actual form is HARMONICALLY BOUNDED.

(6)

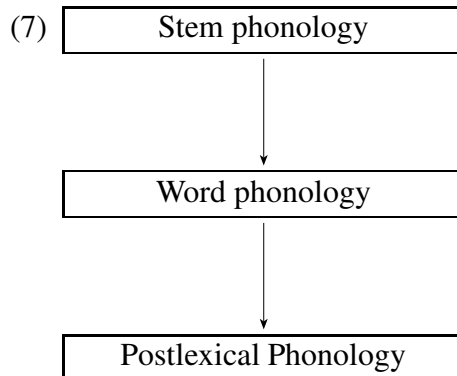
Input	Stage 2: promotion of REDUCE (wrong!)					
	Output	REDUCE	AGREE(Front)	IDENT(Hi)	*ü, ö	IDENT(Front)
uCi	uCi	*	*		*	
	üCi	*			*	*
	uCe			*		
	üCe			*	*	*

The bottom line is that Saussure’s Firewall has no place in constraint-based theories such as OT. This is arguably an advantage because, as noted above, it is stipulative, kills structural explanations of sound change, and even on the descriptive side creates more technical problems than it solves.

3 Stratal OT

Stratal OT phonology provides a more articulated theory than parallel OT in that it incorporates Lexical Phonology’s stratal organization (level-ordering) to OT’s parallelism of constraint interaction (Booij 1996, 1997, Orgun 1996, Kiparsky 2000; for diachronic phonology, see especially Bermúdez-Otero 1999, 2006a, 2006b, Bermúdez-Otero and Hogg 2003). Stratal OT does not in principle banish predictable feature values from lexical representations. Rather, it claims that lexical representations are determined by best satisfaction of the lexical phonological constraints. They will include such redundant feature values as those lexical constraints may assign. For this reason they can accommodate quasi-phonemes.

For Stratal OT, the grammar is a hierarchy of serially related modules, each a parallel constraint system of the classical OT type (without Output-Output constraints, Sympathy constraints, Lexical Conservatism constraints, Base-Reduplication constraints, Turbidity, Targeted constraints, or any other added transderivational devices).



As in Lexical Phonology and Morphology, the Stratal OT levels are morphological as well as phonological subsystems, which form a hierarchy of domains: stems, words, phrases. A constraint system of level $n+1$ may differ in ranking from a constraint system of level n by promotion of

constraints to undominated status. Each is governed by a (parallel) constraint system, but they interface serially. The interaction of constraints is determined by the intrinsic relation of the levels. A constraint at level n is visible to a constraint at level m iff $n \leq m$. Opacity reduces to constraint masking, and “cyclic” effects reduce to ordinary faithfulness: bigger constructions inherit the phonological properties from the smaller constructions they contain, in so far as compatible with the applicable constraints.

Postlexical processes may be restricted to certain prosodic domains, of which the smallest is the CLITIC GROUP, and the larger ones are the PROSODIC PHRASE, the INTONATION GROUP, and perhaps others (Inkelas & Zec 1990). Lexical processes apply to stems (level 1) and prosodic words (level 2).

If we reconstruct quasi-phonemes in Stratal OT as lexically specified but distributionally predictable phonological segment types, we get an interesting additional prediction. In Stratal OT, lexical representations are specified by the word-level constraint system. This entails that quasi-phonemes are elements whose distribution is governed by or relevant to at least one lexical constraint, therefore within the domain of a prosodic word. The same elements may of course also figure in postlexical constraints.⁷

That leads directly to a solution for the secondary split problem. Processes become phonologized when they become applicable to the lexical phonology — formally, when the constraints that drive them are promoted over the antagonistic faithfulness constraints in the lexical constraint system. At that point their outputs become quasi-phonemes, understood as “lexical allophones”. The effect of this promotion is that they assign categorical feature values, that their distribution is determined by constraints that operate within the word domain, and that in virtue of these very facts they are perceptually salient in the sense stated above. Other than the fact that “real” phonemes have an at least partly unpredictable distribution, there is no basic difference between quasi-phonemes and ordinary phonemes, on this view.

The promotion of constraint rankings from the postlexical phonology into the lexical phonology does not mean that those rankings necessarily cease to apply postlexically. The process is, in fact, the generalization of new constraint rankings from the postlexical phonology, where they are first introduced as sound changes, into the lexical (word-level and ultimately stem-level) phonology. The cause of this spread of constraint rankings, I conjecture, is a preference of learners for assigning structure as early as possible. That is, there is a bias in acquisition in favor of locating information in the lexicon.

Although the phonologization of a process in this sense is compatible with its continued postlexical operation, the next step is typically disappearance of its postlexical reflexes — formally, by the promotion of antagonistic faithfulness constraints in the postlexical phonology. Once this happens, there is unambiguous evidence for phonologization, in that the process ceases to apply across word boundaries, its output is strictly categorical, and it is perceptually salient.

In the final act of this phonologization scenario, the potential contrasting quasi-phonemes become overtly manifested. This can happen either when a sound change (the promotion of a con-

⁷For example, in Russian [i] and [i̯] play a role in the lexical phonology, but [i̯] becomes [i] after a velar consonant across a word boundary within a clitic group or phonological phrase. See Rubach 2000, Blumenfeld 2001, Padgett 2003 for discussion of this interesting case.

straint in the postlexical phonology) renders their conditioning environment opaque (this is so-called secondary split), or when new lexical entries from borrowing or other sources exploit them. On this understanding, the rise of phonological contrasts is analogous to the rise of phonological opacity by constraint masking.

Returning to umlaut, we can now offer an analysis of the phonemicization of front rounded vowels. The original grammar has a uniform ranking, with the derivations of (4), in all phonological strata.

As a sound change, umlaut is the acquisition of the constraint ranking (5) in the postlexical phonology. The vowels *ü, ö* (and *æ*, if that is the output of umlaut at this point) are in complementary distribution with *u, o, a*.

In the second phase of the change, the ranking (5) enters the word phonology. At that point, the umlaut vowels become quasi-phonemes, present in lexical representations and constituting inputs to the postlexical phonology. Since lexical umlaut at first applies in a subset of the contexts in which postlexical umlaut applies, this is initially a covert change. It becomes overtly detectable at the latest in the next phase, when back vowels are restored before clitics with *-i-* in clitic configurations (see (9) below), while umlaut continues to apply within the phonological word. Formally, this means that IDENT(Back) is promoted in the postlexical phonology but remains dominated by umlaut in the lexical phonology. The umlaut vowels are not yet overtly contrastive.

In the third phase, another sound change affects the umlaut-triggering *i, j* in such a way as to causes the conditioning of umlaut to become opaque. Let us continue to assume that this happens through the promotion of REDUCE in the postlexical phonology. Lexical umlaut vowels are unaffected, both phonetically and phonologically. The change in the postlexical phonology that masks the context of umlaut does, however, cause them to change from *covertly* contrastive to *overtly* contrastive elements at this point. In principle, they might also become overtly contrastive through the acquisition of any lexical item with an umlaut vowel in a non-umlauting context, whether through borrowing, onomatopoeia, or word-formation, along the lines of the Russian example cited above).

In Old High German, this final phase of the change is reached when postlexical vowel reduction (by promoted REDUCE), *applying to the output of the word-level phonology*, produces minimal contrasts such as *uCe* and *üCe*:

(8)

Input	Overt phonologization: postlexical Vowel Reduction					
	Output	REDUCE	AGREE(Front)	IDENT(Hi)	IDENT(Front)	*ü, *ö
üCi	uCi	*	*		*	
	üCi	*				*
	uCe			*	*	
	☞ üCe			*		*
uCe	uCi	*	*	*		
	üCi	*		*	*	*
	☞ uCe					
	üCe				*	*

Although the postlexical promotion of REDUCE renders the conditioning of umlaut opaque, the

lexical umlaut vowels themselves are retained. They just become overtly contrastive elements at this point.

Whereas Saussure's Firewall prisms apart sound change and phonology and fences them off into separate worlds assigned to distinct fields of inquiry, this alternative explains phonologization through the internal stratification of phonology into a lexical and a postlexical component. And that stratal organization is independently motivated by rich evidence, including cyclic (paradigmatic) effects and phonological opacity. In fact, secondary split is just the historical counterpart of opacity, and Stratal OT provides the same solution to both.

This theory predicts that any phonologization process will proceed in three overt stages. All of them can be documented for umlaut in Old High German. In Otfried's dialect, which we can take here to represent the earliest stage, after the sound change enters the language, umlaut was postlexical, and hence crossed lexical word boundaries, applying within clitic groups.

- (9) a. /mag iz/ → *meg iz* 'may it'
b. /drank ih/ → *drenc ih* 'I drank'
c. /gab ima/ → *geb ima* 'gave him'
d. /giraH inan/ → *gireh inan* 'avenged them'

In early OHG, umlaut became a lexical process, and ceased to apply across word boundaries, but was still transparently conditioned within the lexical word. The umlaut vowels were now quasi-phonemes. In the third stage, they became overtly contrastive as a result of sound changes that rendered their conditioning environments opaque.

The theory also predicts that our three criteria for quasi-phonemes should be satisfied at the second stage. As far as it is possible to tell, this is the case. The first criterion is certainly satisfied, for umlaut at that stage became restricted to applying inside lexical words. The second criterion is also satisfied: umlaut vowels must have been more salient exponents of vowel frontness than their triggers, at least in the normal cases where the umlaut vowels are stressed and the context is unstressed. The third, categoriality, is hardest to verify. The vowels *ii*, *ö* began to be written only late, because the Latin alphabet had no letters for them, but the umlaut of *a* was written *e* already at the second stage, that is, well *before* the reduction of *-i* to *-e* that (on the structuralist view) caused it to become phonemic. This could be taken as an indication that they were perceived as categorially distinct from *a* at stage 2, i.e. prior to the point at which the structuralist theory of phonologization claims that they became phonemic.

An attractive facet of the proposed Stratal OT approach to phonologization is that it establishes an inherent causal connection between the loss of the triggering context and the rise of the new phoneme, rather than a merely accidental one. In the case at hand, the progressive reduction of unstressed syllables is a *cause* of the concurrent rise of the umlaut vowels to quasi-phonemic status. Even before the reduction leads to neutralization, it causes the primary cue to shift from the unstressed vowels to the stressed vowels, which is a precondition for the latter to become quasi-phonemes.

We now have a solution to the puzzle why English palatalization of *k-* and *g-* does not become contrastive when vowels that condition it change their backness value. The answer is that the

palatalization process is postlexical. We know that because it is gradient rather than categorical, because it is not salient, and because it applies in close contact across word boundaries, e.g. *sock it* vs. *sock us*. Stratal OT predicts that under these circumstances it cannot become phonemic by secondary split. Therefore, vowel fronting and backing does not result in a contrast between front and back *k*.

Crucially, Stratal OT departs from Lexical Phonology by giving up structure-preservation (“Strata, yes, structure-preservation, no”, as the slogan of Roca 2005 has it). To put it another way, Stratal OT severs the structuralist link between CONTRASTIVENESS (unpredictable distribution), a *structural* notion, and DISTINCTIVENESS, a *perceptual* notion.⁸ Phonemes are contrastive and distinctive, allophones are non-contrastive and non-distinctive. The other two combinations are the surprising ones. Quasi-phonemes are non-contrastive but distinctive — that is, they are predictable but perceptually salient. The fourth logically possible case, contrastive but nondistinctive elements, exists as well. These are NEAR-MERGERS (Labov 1994, Ch. 12), as when a speaker reliably produces near-merged sounds slightly differently, but cannot distinguish between them, in the speech of other such speakers or in her own speech, e.g. *source* and *sauce* in New York. The four cases are shown in (10).

(10)

	contrastive	non-contrastive
distinctive	phonemes	quasi-phonemes
non-distinctive	near-mergers	allophones

The upshot is that while delinking contrastiveness and distinctiveness in a sense preserves the phoneme as a theoretical construct, it does so only by negating the founding intuition behind it.⁹

Finally, Stratal OT also offers a solution to the empirical problems for Saussure’s Firewall that we identified above. It predicts that sound changes will relate transparently to other postlexical processes. This has the three consequences that we cited above as difficulties for Saussure’s Firewall.

First, when conditioned allophones are created in the postlexical constraint system, they will just disappear when their conditioning environments are lost, and no secondary split will occur. In other words, sound changes can bleed existing postlexical processes. That is, they can eliminate some of their former inputs. English velar to palatal assimilation is postlexical, since it is determined by the context across word boundaries (e.g. *sock it* vs. *sock us*). Stratal OT predicts that under these circumstances it cannot become phonemic by secondary split. Therefore, vowel fronting and

⁸This link was axiomatic at least in post-Bloomfieldian American structuralism. Bloomfield himself allowed distinctive sounds to be non-contrastive, for example if they were morphologically predictable, a practice later condemned as “mixing levels”. The Prague school distinction between phonetic and allophonic processes might also be seen as implying the separation of distinctiveness from contrastiveness.

⁹It also calls into question Natural Phonology’s idea that the phoneme is ‘the mental image of a sound’ whose perception makes contrast possible (‘minimal pairs exist because phonemes are perceived as distinct’, Nathan & Donegan, this volume). The point about near mergers is precisely that they give rise to minimal pairs that are not perceived as distinct. Conversely, quasi-phonemes are perceived as distinct even if they do not contrast (Hellberg’s 1978 analysis of Swedish [æ] is particularly instructive on this point)). Decoupling contrastiveness and distinctiveness in Natural Phonology is not trivial.

backing sound changes do not result in a contrast between front and back *k*. While quasi-phonemes survive the loss of their conditioning environment, postlexical allophones disappear.

The second consequence is that a sound change can feed other existing postlexical processes, i.e. add new inputs to them. Consider a language that has obligatory voicing assimilation of obstruents within some postlexical domain, such as the phonological phrase or the phonological word (the clitic group). The prediction is that when sound change creates sequences of obstruents in such a language, voicing assimilation will automatically eliminate them, as in the previously mentioned Old English example *bidest* (> **bidst*) > *bitst*. The parenthesized intermediate form is a “virtual” stage which is not pronounced but forms part of the sound change itself.

Third, sound changes can be blocked just in case their output does not conform to a constraint that holds at the postlexical level.

In order to account for secondary split and neogrammarian exceptionlessness, we do not have to stipulate that the promotion of constraints is limited to the postlexical stratum. Constraints can be reranked at any stratum. Reranking at the word and stem levels simply amounts to another type of change, namely analogy (including LEXICAL DIFFUSION, the extension of a lexical rule to new items, Kiparsky 1995, also Phillips, this volume).

This can be illustrated with the more recent development of umlaut in German. It has split into a stem-level and a word-level process. The word-level process applies to word-based formations made with inflectional suffixes and some productive derivational suffixes, illustrated here by the comparative suffix *-er*. Historically, it is the result of an analogical streamlining of the synchronic umlaut process. It only triggers vowel *fronting*, and only in a syllable *adjacent* to the triggering suffix. Stem-level umlaut, on the other hand, preserves the inherited umlaut process with its historically accreted complexities. It generates, in addition to vowel fronting, the synchronically arbitrary rounding switch of *au* to *oi* (spelled *äu*), as in (11b,c,d,e), and it applies non-locally across a syllabic sonorant, as in (11d,e,f).

(11)	Stem Level	Word Level (Word+Suffix)
a. <i>arm</i> ‘poor’	<i>ärmlich</i> ‘impoverished’	<i>ärmer</i> ‘poorer’
b. <i>blau</i> ‘blue’	<i>bläulich</i> ‘bluish’	<i>blauer</i> ‘bluer’
c. <i>braun</i> ‘brown’	<i>bräunen</i> ‘to brown’	<i>brauner</i> ‘brownier’
d. <i>sauber</i> ‘clean’	<i>säuberlich</i> ‘tidy’	<i>sauberer</i> ‘cleaner’
e. <i>lauter</i> ‘pure’	<i>läutern</i> ‘to purify’	<i>lauterer</i> ‘purer’
f. <i>schwanger</i> ‘pregnant’	<i>beschwängern</i> ‘to impregnate’	<i>schwangerer</i> ‘more pregnant’

Level ordering unifies what superficially look like two distinct umlaut processes in the synchronic phonology, and allows the generalization that German umlaut *never* crosses a syllable. This phonological locality restriction is common to both the word-level and stem-level versions of the process (e.g. *Bubi* → *Bubi-lein*, not **Bübi-lein* “little Bubi”). Because *r* is not syllabified at the stem level (as can be shown on independent grounds), umlaut in cases like *beschwängern* actually does not cross a syllable. Final *r* after *-C* becomes syllabic at the word level, blocking umlaut in word-based formations such as *schwäng[r]-r*.¹⁰

¹⁰It is also part of the explanation for the contrast between umlaut in common adjectives versus lack of umlaut

This later development of umlaut illustrates the next stage of the typical LIFE CYCLE of phonological processes (Hyman 1976, Iverson & Salmons 2009, Roberts 2012, Bermúdez-Otero, this volume). In terms of Stratal OT, the full trajectory begins with gradient variable phonetic implementation processes, which become incorporated into the phonology as featurally discrete constraints, first enforced postlexically without any domain restrictions, and then get restricted to increasingly narrow domains, finally retiring as morphologically or lexically conditioned alternations (Bermúdez-Otero & Trousdale forthcoming). Any phonological theory must account for the robust directionality of the phonological life cycle.¹¹ One obvious hypothesis would be that the drift reflects learners' bias in favor of precompiling the output and restricting constraints to the narrowest possible lexical domain. Such a bias should be eminently testable in childrens' language acquisition; in any case it seems more plausible than the "mystical, pan-generational forces" invoked by Hale, Kissock & Reiss (this volume).

4 Conclusion

Sound change and phonologization pose theoretical problems for structural and generative phonology which are conceptually akin to the ones raised by opacity in synchronic phonological systems. I have argued that Stratal OT provides a parallel solution to both sets of problems, which involves a radical rethinking of the phoneme. A corollary of the move to OT is that the time-honored formulation of sound changes as context-sensitive replacement processes (A → B / C__D), for all its convenience, is misleading in that it fails to represent the way the process and/or its conditioning environment may be motivated or constrained by the language's existing phonological system.

I conclude by briefly mentioning some further predictions that follow directly from the proposed account of phonologization. Intrinsically lexical processes *must* become opaque if sound change masks their conditioning, for sound changes are by hypothesis postlexical, and therefore cannot affect the operation of lexical processes. For example, consider a language where word stress is assigned in the lexical phonology. A sound change which crucially affects the context that determines the place of stress — let us say an apocope process, or resyllabification — necessarily leaves the stress unchanged. Later analogical changes can of course restore the transparency of stress in various ways, but this cannot happen as part of the original process.

Another consequence is that sound changes cannot be conditioned by word boundaries. That this is the case has been argued by Hock (1991: 239) on the grounds that word boundary restrictions in many cases demonstrably originate as phrasal boundaries. That is, at the first stage word boundaries play a role only insofar as they are phonetically manifested by pauses or other audible effects.

For similar reasons, we predict that secondary split should be controlled by the postlexical phonology, e.g. deletion of C in V.CV will result in disyllabic V.V just in case hiatus is tolerated in

in rarer adjectives, including in particular compounded forms of some of the same adjectives: *klüg+er* "smarter", *alt+klug+er* "more precocious", *kält+est+e* "coldest", *eis+kalt+est+e* "ice-coldest". Similarly, the consonantal alternation in *näch+ste* "nearest" (from *nah* "near") does not apply in the compound *haut+nah+st+e* "nearest to the skin".

¹¹In rare cases, a single step is reversed, but the whole trajectory never runs backwards: as far as we know no morphological alternation has ever undergone a stepwise domain widening and ended up as a phonetic implementation process.

the postlexical phonology (i.e. if *ONSET is dominated by syllabic faithfulness in that component). If the constraint ranking in postlexical phonology favors consonant epenthesis over hiatus, it will yield a disyllabic *VCV* sequence with an inserted consonant. If the postlexical phonology bars all derivations of a disyllabic output, but permits vowel length, the deletion will yield monosyllabic *VV*.

A constraint-based approach also makes immediate sense of the structure-preserving tendency that has been noted for sound change, that features which fit into the system are more easily lexicalized. For example, the generalization that redundant tones tend to be phonologized in systems which already have tonal contrasts (Kiparsky 1995) follows, since in such languages the constraints of the lexical phonology that admit distinctive tonemes are also likely to admit non-distinctive quasi-tonemes.

Stratal OT was devised as a theory of synchronic phonology. But it makes predictions about sound change that can be subjected to empirical scrutiny. Thus linguistic change gives us a window not only on the grammars of individual languages, but on the principles that underlie all of them.

Bibliography

- BENEDIKTSON, HREINN. 1982. Nordic umlaut and breaking: Thirty years of research (1951–1980). *Nordic Journal of Linguistics* 5: 1-60.
- BERMÚDEZ-OTERO, RICARDO 1999. Constraint interaction in language change: quantity in English and Germanic. Ph.D. Dissertation, University of Manchester.
- BERMÚDEZ-OTERO, RICARDO, AND RICHARD M. HOGG. 2003. The actuation problem in Optimality Theory: Phonologisation, rule inversion, and rule loss. In D. Eric Holt (ed.), *Optimality Theory and language change*. Dordrecht: Kluwer.
- BERMÚDEZ-OTERO, RICARDO 2006a. Diachronic phonology. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*. Cambridge: CUP.
<http://myweb.tiscali.co.uk/bermudez/handbook.pdf>
- BERMÚDEZ-OTERO, RICARDO 2006b. Phonological change in Optimality Theory. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn., vol. 9, 497-505. Oxford: Elsevier. <http://myweb.tiscali.co.uk/bermudez/encyclopedia.pdf>
- BERMÚDEZ-OTERO, RICARDO. Diachronic origins of opacity and incomplete neutralization. This volume.
- BERMÚDEZ-OTERO, RICARDO & GRAEME TROUSDALE (forthcoming). Cycles and continua: on unidirectionality and gradualness in language change. In Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *Handbook on the history of English: rethinking and extending approaches and methods*. Oxford: Oxford University Press.
- BLEVINS, JULIETTE. 2004. *Evolutionary Phonology*. Cambridge: CUP.
- BLUMENFELD, LEV. 2001. Russian palatalization in Stratal OT. MS, Stanford University.
- BOOIJ, GEERT. 1996. Lexical Phonology and the derivational residue. In Durand, Jacques & Bernard Laks (eds.) *Current Trends in Phonology: Models and Methods*. European Studies Research Institute and University of Salford.
- BOOIJ, GEERT. 1997. Non-derivational phonology meets Lexical Phonology. In Iggy Roca (ed.) *Derivations and constraints in the lexicon*. Oxford: Clarendon Press.
- BRAUNE, WILHELM. 2004. *Althochdeutsche Grammatik I: Laut- und Formenlehre*.¹⁵ Ed. Ingo Reiffenstein. Tübingen: Niemeyer.
- FLEMMING, EDWARD. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7-44.
- HALE, MARK, MADELYN KISSOCK & CHARLES REISS . An I-Language approach to phonologization and lexification. This volume.
- HELLBERG, STAFFAN. 1978. Unnatural phonology. *Journal of Linguistics* 14: 157–177.
- HOCK, HANS HENRICH. 1991. *Principles of historical linguistics*. Berlin: Walter de Gruyter.
- HOENIGSWALD, H. 1965. *Language change and linguistic reconstruction*. Chicago: University of Chicago Press.
- HYMAN, LARRY. 1976. Phonologization. In *Linguistic Studies offered to Joseph Greenberg*, A. Juillard (ed.), Anma Libri: Saratoga, CA, 2:407-418.
- INKELAS, SHARON, & DRAGA ZEC. (1990). *The Phonology-Syntax Connection*. Chicago: University of Chicago Press.

- IVERSON, GREGORY K., AND JOSEPH C. SALMONS. 2009. 2009. Naturalness and the lifecycle of sound change. *On Inflection: In Memory of Wolfgang U. Wurzel*, ed. by Patrick Steinkrüger & Manfred Krifka, 89-105. Berlin: Mouton de Gruyter.
- JAKOBSON, ROMAN, GUNNAR FANT, AND MORRIS HALLE. 1952. *Preliminaries to speech analysis*. Cambridge: MIT Press.
- JANDA, RICHARD D. 2003. "Phonologization" as the start of dephonicization — or, on sound change and its aftermath: of extension, lexicalization, and morphologization. In Joseph, Brian D. and Richard D. Janda (eds.), *The handbook of historical linguistics*, p. 401-422. Oxford: Blackwell.
- JASANOFF, JAY. 2003. Plus ça change... Lachmann's Law in Latin. *Harvard Working Papers in Linguistics* 8.
- KING, ROBERT D. 1971. Syncope and Old Icelandic *i*-umlaut. *Arkiv för Nordisk Filologi* 86: 1-18.
- KING, ROBERT D. 1973. Rule insertion. *Language* 49: 551-578.
- KIPARSKY, PAUL. 1995. The Phonological Basis of Sound Change. In John Goldsmith (ed.) *Handbook of Generative Phonology*. Oxford, Blackwell. Reprinted in Richard Janda and Brian Joseph (eds.) *Handbook of Historical Linguistics*. Oxford, Blackwell, 2003.
- KIPARSKY, PAUL. 1998. Sievers' Law as prosodic optimization. In Jay Jasanoff, H. Craig Melchert, and Lisi Oliver (edd.), *Mír Curad: Studies in honor of Calvert Watkins*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- KIPARSKY, PAUL. 2000. Opacity and cyclicity. *The Linguistic Review* 17:351-367, 2000.
- KIPARSKY, PAUL. 2006. Amphichronic linguistics vs. Evolutionary Phonology. *Theoretical Linguistics*, 32: 217-236.
- KIPARSKY, PAUL. To appear. Grammaticalization as optimization. In Diane Jonas (ed.), *Proceedings of the 8th Diachronic Generative Syntax Conference*.
- KIPARSKY, VALENTIN. 1932. Johdatusta fonologiaan. *Virittäjä* 36: 230-250.
- KORHONEN, MIKKO. 1969. Die Entwicklung der morphologischen Methode im Lappischen. *Finnisch-Ugrische Forschungen* 37: 203-62.
- LABOV, WILLIAM. 1994. *Principles of language change*, Vol. 1. Oxford: Blackwell.
- LIBERMAN, ANATOLY. 1991. Phonologization in Germanic: Umlaut and vowel shifts. In Elmer H. Antonsen and Hans Hock (eds.) *Stæfcræft. Studies in Germanic linguistics: Select papers from the 1st and 2nd symposium on Germanic linguistics*. Amsterdam: Benjamins.
- ORGUN, ORHAN. 1996. Sign-based phonology and morphology, with special attention to optimality. Ph.D. dissertation. University of California, Berkeley.
- PADGETT, JAYE. 2003. Contrast and post-velar fronting in Russian. *Natural Language & Linguistic Theory* 21: 39-87.
- ROBERTS, PHILIP. 2012. Towards a Computer Model of the Historical Phonology and Morphology of Latin. Ph.D. Dissertation, Oxford University.
- ROCA, IGGY. 2005. Strata, yes, structure-preservation, no. In Geerts, Twan, Ivo van Ginneken, and Haike Jacobs (eds.) *Romance languages and linguistic theory 2003*. Amsterdam: Benjamins.

- RUBACH, JERZY. 2000. Backness switch in Russian. *Phonology* 17, 39-64.
- DE SAUSSURE, FERDINAND. 1914. *Cours de linguistique générale*. Transl. Wade Baskin, *Course in General Linguistics*. New York: Philosophical Library, 1959.
- DE SAUSSURE, FERDINAND. 1993. *Saussure's Third Course of Lectures on General Linguistics (1910-1911)*, ed. and transl. E. Kimatsu and Roy Harris, Oxford: Pergamon.
- TWADDELL, W. FREEMAN. 1938. A note on OHG umlaut. *Monatshefte für deutschen Unterricht* 30.177-181.