1 Derivations and levels of representation

Two intimately related concepts in linguistics are level and derivation. Generative phonology recognizes a minimum of two levels of representation, generally known as the underlying and surface levels. The derivation is the mapping from underlying to surface representation. So far there is nothing uniquely generative about this conception. Levels and mappings between levels are known from structuralism, so it is important to understand how structuralist and generative notions of language differ in order to get an adequate grasp of both of these concepts.

At its broadest, the level in structuralist analysis refers to a way of describing a linguistic expression from a particular point of view. A sentence like *the Celts wore kilts*, for example, can be described from a syntactic perspective [*the Celts* [*wore* [*kilts*]]], morphological {determiner N+plural V+past N+plural}, and phonetic [ðə 'kelts wɔː 'kɪlts]. Each of these descriptions is autonomous, with its own vocabulary and principles of well-formedness. In describing the sound pattern of a language, structuralism minimally invoked two levels of description, termed phonemic and phonetic (a.k.a. allophonic). The phonetic level supplies a more or less detailed transcription of an utterance understood as an individual speech event at a particular time an place, or an equivalence class of such speech events. A narrow transcription of our sentence might be [ðə kʰɛɫtʰs wɔː kʰɪlts]. The phonemic level abstracts away from non-contrastive variation in the speech signal such as post-aspiration, the velarization of laterals, or precise shades of vowel quality. A phonemic transcription of the same sentence might therefore be /ðə kɛltswana: kɪlts/. The end result of phonemic analysis is a set of autonomous phonemes along with a statement of their allophones and the contexts in which they appear. Beyond this, it is possible to distinguish three different philosophical positions on the phoneme that we can designate by the terms psychological realism or mentalism, physical realism and nominalism. Linguists such as Edward Sapir (1925), Nikolai Trubetzkoy, and Baudouin de Courtenay (Anderson, 1985, p. 290) believed that the phoneme could be understood as the mental image of a speech sound. The second position, associated with Leonard Bloomfield (1933) and Daniel Jones (1962), was that the phoneme was the acoustically invariant part of its allophones. The third position, first articulated by Twaddell (1935), was that the phoneme didn’t refer to an objective reality at all; phonemes are merely analytical conveniences posited by the linguist so as to be able to describe the contrasts of a particular language. It was this view that came to serve as the basis for American Structuralism, which was the immediate context for the development of generative phonology.

The shift away from these structuralist conceptions began in the 1920s as distinctive feature theory began to take shape with the collaboration of Nikolai Trubetzkoy and Roman Jakobson in

I would like to thank Martin Krämer, and two anonymous reviewers for their invaluable feedback.
Their work on neutralization ended up breaking with Saussure’s purely relational conception of the phonological system as a system of unanalyzable phonemes (Saussure, 1995 [1916]). The classic example, from German, is the neutralization of the phonemic distinction between /t/ and /d/ to [t] in word-final position, as illustrated by alternations such as Rad [rät] ‘wheel’ vs. the corresponding dative form Rade [räta] (cf. Rat [rät] ‘counsel’, with dative Rate [räta]). Because they exclusively alternated, /t/ and /d/ were shown to be clearly more closely related to each other, or ‘correlated’, than with other phonemes. Pairs of correlating phonemes were grouped into archiphonemes each with a marked and an unmarked member distinguished by a principum divisionis, in this case voicing. Neutralization also had a particular direction: it always entailed the loss of the principum divisionis so the result of the neutralization was always the unmarked member of the pair. This complex of ideas represents the beginning of reasoning phonologically in terms of natural classes, distinctive features and markedness, as well as dynamic rules.

In the years leading up to the inception of the generative paradigm in phonology, structuralist phonological analysis recognized a further morphophonemic level (Swadesh and Voegelin, 1939). Just as the phonemic level abstracted away from variation due to phonetic context, the morphophonemic level was conceived to abstract away from alternations conditioned by the phonological or morphological context to permit a unique representation for each morpheme. For example, in order to handle the difference between non-alternating singular/plural pairs such as hive∼hives, fife∼fifes and alternating pairs like wife∼wives, a morphophoneme /fM/ was posited to cover cases where the phonemes /f/ and /v/ alternated. This allowed the formulation of statements specifying how morphophonemes mapped onto phonemes in particular morphophonological contexts. While recognizing morphophonemes as a convenient descriptive device, Hockett (1961) noted that the morphophonemic level was not itself autonomous, but was “produced […] by conflicts between the independent workings of grammar and of phonology” (p. 50). According to the structuralist conception, it was the phonemic level of description that was autonomous. With the advent of generative phonology, however, this understanding ended up being turned on its head. Beginning with Chomsky (1951), it was the morphophonemic level, or something approximating it (now generally known as the ‘underlying level’), that came to be seen as autonomous, and the autonomous phonemic level of structuralism was abandoned. This was part and parcel of a fundamental shift in the understanding of what linguistic symbols represented. In contrast to the structuralist levels of analysis, the underlying and surface levels in generative phonology are both made up of the same discrete symbolic units (distinctive features, segments, and so on). These units purport to be psychologically real and have intrinsic phonetic content, as opposed to abstract phonetically uninterpretable morphophonemes.1 The underlying and surface levels occupy unique positions in the grammar as a whole. The underlying form is what is encoded in the lexicon. The surface level, on the other hand, serves as an interface to the articulatory and auditory systems.

The reason for the paradigm shift can be traced to two crucial demonstrations that positing an autonomous phonemic level resulted in missing linguistically significant generalizations in some cases. Halle (1959) showed first that an autonomous phonemic level entailed duplicating generalizations when the effect of a rule was neutralizing in some cases and non-neutralizing in others. Russian obstruents generally contrast in voicing, as shown by minimal pairs such as /plʲi/ ‘dust-GEN.SG’ vs. /blʲi/ ‘be-PAST.PL’, and /got/ ‘year’ vs. /kot/ ‘tomcat’. An obstruent assimilates in voicing to a following obstruent, e.g. /k domu/→[gdomu] ‘towards the house’, /spas bi/→[spazbi] ‘would have saved’, /ot'ëts bi/→[at'ëdzbi] ‘father was’, /dot' biilá/→[dodzbi]á] ‘daughter was’. Because Russian has a voicing contrast in obstruents, obstruent voicing assimilation is generally neutralizing, but in
the case of /ts/ and /tʃ/, which lack voiced counterparts, it is not. On a structuralist analysis, the neutralizing part of the rule would have to be stated as a mapping from the morphophonemic to the phonemic level, while the non-neutralizing part would have to be stated as a mapping from the phonemic to the allophonic level.

A few years later, Chomsky (1964) showed that phonological rules are not simply directed mappings from an underlying to a surface representation. Phonological rules interact and must therefore be linearly ordered, and failing to recognize this can lead to positing false contrasts. In Canadian English (Joos, 1942), /ɔ̄y/ is raised to [ʌy] before voiceless stops, and /t d/ are realized as a flap intervocally. Because /t/ is an obstruent stop, the contexts of these rules overlap, which makes it necessary to assign priority of application to one rule or the other. As a word like writer [rʌyɾɚ] shows, raising must apply before flapping. Applying the rules in the opposite order would erroneously predict [rɑ̄yɾɚ] which, in the dialect in question, corresponds to the pronunciation of a different word, rider. The crucial ordering of these two rules thus results in surface minimal pairs which, if used as the sole basis for determining contrast, can overdetermine the underlying inventory. Even though the diphthongs [ɔ̄y] and [ʌy] contrast exclusively before [ɾ], the structuralist procedure for discovering phonemes, which, precisely, proceeds on the basis of (surface) minimal pairs like [rʌyɾɚ] vs. [rɔyɾɚ], would force us to posit two phonemes, /ɔ̄y/ and /ʌy/. What the taxonomic account fails to explain is how it comes about that [ʌy] is distinct from [ɔ̄y] preceding exactly those flaps that alternate with a voiceless obstruent /t/.

Furthermore, the processes involved in deriving the output are not restricted to adding information, such as velarization of an underlying /l/, spreading tone to tonally unspecified vowels, building syllables over unparsed segments, and so on. Processes may also carry out destructive alterations of the underlying representation, for example by deleting segments or changing feature specifications. All neutralization rules are destructive by definition. This is a crucial difference between derivational and non-derivational theories.

These developments led to a shift of interest away from levels and their properties to rules and the way they interact in derivations. Goldsmith (1993, 25), writing of developments in the late 1960s goes so far as to say that “the notion of derivation changed, in many linguists’ perception from being an account of the fundamental problem of levels in linguistics to being the essence itself of a linguistic analysis”. In early generative phonology, the underlying and surface representations exert no influence on the outcome of the derivation, except in the sense that the underlying representation provides the input to the first rule. As the derivation progresses, the system retains no memory of earlier states, nor is it guided by representations of desirable future states, or functional goals. What ultimately replaced the level-based view was “a dynamic model of linguistic analysis, in terms of which one representation is successively changed into another in a sequence that in its entirety is the account of the expression in question”. Nevertheless, throughout the history of the field’s development, phonologists have repeatedly come back to the the role and properties of levels in their theories. This interest manifests itself in a variety of different ways that always reflect the formal technology of the time. We shall see several examples of this in the following pages.

By way of rounding off this section, let us take a closer look at the question which phonological theories should be considered derivational. Derivational theories make use of at least two consubstantial levels (underlying and surface) and potentially destructive processes that change the one into the other. This definition rules out monostratal theories, which only recognize a single level of representation. Declarative Phonology (Scobbie, 1991, 1993; Scobbie et al., 1996) is restricted to stating generalizations over surface forms, and is for this reason neutralization poses a considerable
challenge. Constraints in Declarative Phonology may be viewed as either structure-checking filters, which rule representations as either grammatical or ungrammatical, or as purely structure-building operations that combine information with information already contained in the input. Any grammar consisting solely of feature-filling processes has a monostratal interpretation, and vice versa. The original version of Optimality Theory (OT; Prince and Smolensky, 2004 [1993]; McCarthy and Prince, 1993) may also be understood as monostratal, since it assumes the Principle of Containment, according to which nothing is literally deleted or changed in the mapping from input to candidate output. Each output candidate simply offers an alternative parse of the input; incorporating earlier ideas (see esp. Ito, 1989), inaudible material simply remained syllabically unparsed rather than being deleted outright, while epenthesis is understood as the insertion of segmentally unfilled nuclei. OT with Correspondence Theory (McCarthy and Prince, 1995, 1999), however, since it countenances feature-changing processes, is derivational. The main difference between rule-based phonology and OT with Correspondence is whether linguistically significant generalizations reside in constraints on representations or in the rules that map one representation onto another.

Also plainly non-derivational are theories where the levels are made up of different kinds of thing. An example is Firthian prosodic analysis (Firth, 1948; Ogden, 1993), where the discrete units of symbolic analysis, known as ‘prosodies’, are mapped onto a continuous parametric phonetic representation.

Derivational is often popularly understood as synonymous with serial. However, as we shall see examples of below, processes may also apply simultaneously or in parallel in a derivational theory. Finally, there is the perennial question to what extent linguistically significant generalizations may be said to inhere in levels of representation or the mapping between levels. Early rule-based theory and Reiss’ (2008) rule-only approach lays this burden entirely on the derivation. In monostratal theories and in Classical OT, on the other hand, generalizations reside exclusively in constraints on the surface representation. Intermediate positions are possible. Level-based theories such as Two-Level Morphology (Koskenniemi, 1983; Karttunen, 1993), Harmonic Phonology (Goldsmith, 1993), and Cognitive Phonology (Lakoff, 1993) posit constraints both on privileged levels of representations (the number of which varies depending on the proposal) and on mappings between levels. These are construed as essentially static correspondences, lacking the interactions that characterize the derivations of rule-based theory. Although there are important points of similarity between OT and level-based theories (OT’s output-orientedness), OT with Correspondence Theory works with processes. Late rule-based theory includes constraints on input and output, and later developments in OT may make additional appeal to derivational notions.

In the next two sections we will look at the interactions constitutive of derivations, and how phonologists have tried to argue for limiting them, often at the cost of descriptive adequacy. Section 2 examines the interaction of rules and processes and gives examples of their effects on surface representation. This section also discusses theories of rule and process ordering. Section 3 considers phonological and non-phonological alternatives for dealing with the empirical residue in theories excluding certain types of interaction. Section 4 presents the main conclusions.

2 Derivations in phonological theory

In this section, we look at theories of rule and process ordering. Section 2.1 introduces extrinsic ordering, where rules apply in a strict linear sequence determined by the grammar. Extrinsic ordering
allows rules to interact in potentially complex ways. Section 2.2 discusses the implications of rule interaction for the abstractness of underlying representations. In Section 2.3 we examine the issue whether certain types of interaction are more natural than others. Section 2.5 looks at theories that eschew the extrinsic ordering of rules and processes, including OT in its parallel and harmonic serial implementations. The last two sections look at intrinsic ordering. Section 2.6 looks at within-cycle intrinsic ordering (the Elsewhere Principle), and Section 2.7 addresses the way in which the construction of morphological and syntactic constituents interacts with the phonology.

2.1 Extrinsic rule ordering and transient rules

In SPE, derivations are viewed as a series of ordered rewrite rules of the form A→B/C__D, ‘A is rewritten as B in the environment C__D (between C and D)’ (Chomsky and Halle, 1968; Bromberger and Halle, 1989). Rules specify a process or structural change (A→B) to a focus (here, A) in a given environment (here, C__D). Some rules are context-free, specifying only a focus and a structural change, e.g. A→B. The union of focus and any environment is known as the structural description. The structural description for our hypothetical rule is thus /CAD/. Since rules apply in a strict linear order, each rule only has one chance to apply (or, in a model that recognizes cycles, one chance in a given cycle). To use a term from Chafe (1968), each rule is transient. During the 1970s and 80s, this rule-based approach was elaborated with additional formal devices, including inventories, Morpheme Structure Conditions (MSCs), filters and constraints, as well as highly articulated theories of non-linear representations (Autosegmental Phonology, Feature Geometry and Metrical Phonology). A representative example of how rule-based phonology had come to evolve in the 1990s is Kenstowicz (1994).

Given two mutually affecting rules, P and Q, either P creates strings that satisfy the structural description of Q, or P destroys strings that satisfy the structural description of Q. We can now define the following two-rule interactions in (1).5 The symbol > is to be read ‘precedes’.

\[(1)\]
\[a. \text{ If } P \text{ creates strings that satisfy the structural description of } Q, \text{ and both } P \text{ and } Q \text{ apply, then } P > Q \text{ and we say } P \text{ feeds } Q.\]
\[b. \text{ If } P \text{ destroys strings that satisfy the structural description of } Q, \text{ and } P \text{ applies and } Q \text{ does not apply, then } P > Q \text{ and we say } P \text{ bleeds } Q.\]
\[c. \text{ If } P \text{ creates strings that satisfy the structural description of } Q, \text{ and } P \text{ applies and } Q \text{ does not apply, then } Q > P \text{ and we say } P \text{ counterfeeds (fails to feed) } Q.\]
\[d. \text{ If } P \text{ destroys strings that satisfy the structural description of } Q, \text{ and both } P \text{ and } Q \text{ apply, then } Q > P \text{ and we say } P \text{ counterbleeds (fails to bleed) } Q.\]

In extrinsic ordering theories, the derivation may create intermediate representations. Intermediate representations occur in feeding and counterbleeding interactions. P creates an intermediate representation to which Q then applies. Some phonologists may speak of intermediate representations as ‘levels’, but unlike the underlying and surface levels, intermediate representations are mere epiphenomena of sequential rule application. This usage may, however, be taken as supporting Goldsmith’s general point that a strong emphasis on derivations at certain points in the historical development of the field may obscure the importance of levels.6

Let us now illustrate each of the rule interactions in (1) in turn. The data in (2) show alternations between the plural and neuter singular forms of adjectives in Norwegian (cf. Kristoffersen, 2000, p. 109). The plural is marked by a suffix {-a}, the neuter singular by {-t}. In a triliteral cluster of
Son ⊏ Obs ⊏ Obs, the first obstruent is deleted (Cluster Simplification), as shown in the neuter singular forms for ‘dull’ and ‘glossy’. The neuter singular form of ‘clever’ shows a second alternation. A coronal consonant coalesces with a preceding /ɾ/ to yield the corresponding retroflex. Here, the sequence /rt/ coalesces to [ʈ]. As the neuter singular forms of ‘sharp’ and ‘dark’ show, Cluster Simplification must precede, and feed, Coalescence. UR = underlying representation; SR = surface representation.

(2)  
<table>
<thead>
<tr>
<th>UR</th>
<th>PL</th>
<th>N.SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tyn/</td>
<td>tynə</td>
<td>tynt</td>
</tr>
<tr>
<td>/kʋɑlmt/</td>
<td>kʋɑlmtə</td>
<td>‘nauseous’</td>
</tr>
<tr>
<td>/lʉːʈ/</td>
<td>lʉːʈə</td>
<td>‘clever’</td>
</tr>
<tr>
<td>/sʋɑːʈ/</td>
<td>sʋɑːʈə</td>
<td>‘huge’</td>
</tr>
<tr>
<td>/blaŋt/</td>
<td>blaŋktə</td>
<td>‘glossy’</td>
</tr>
<tr>
<td>/skɑʈ/</td>
<td>skɑʈə</td>
<td>‘sharp’</td>
</tr>
<tr>
<td>/mœʈ/</td>
<td>mœʈə</td>
<td>‘dark’</td>
</tr>
</tbody>
</table>

Sample derivations for the neuter singular forms of ‘clever’, ‘glossy’, and ‘sharp’ are given in (3). Where the structural description is not met at the relevant point in the derivation, we write n/a (not applicable).

(3)  
<table>
<thead>
<tr>
<th>UR</th>
<th>/lʋt+t/</th>
<th>/blaŋk+t/</th>
<th>/skɑt+t/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Simplification</td>
<td>n/a</td>
<td>blaŋt</td>
<td>skɑt</td>
</tr>
<tr>
<td>Coalescence</td>
<td>lʋt</td>
<td>n/a</td>
<td>skɑt</td>
</tr>
<tr>
<td>SR</td>
<td>[lʋt]</td>
<td>[blaŋt]</td>
<td>[skɑt]</td>
</tr>
</tbody>
</table>

An example of bleeding is provided by Lithuanian (Kenstowicz and Kisseberth 1973, 2; see also Ambrazas 1997, 75).

(4)  
<table>
<thead>
<tr>
<th>UR</th>
<th>INF</th>
<th>1SG.FUT</th>
<th>IMP.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>/gerti/</td>
<td>gersiu</td>
<td>gerkite</td>
<td>‘drink’</td>
</tr>
<tr>
<td>/kasti/</td>
<td>kasiu</td>
<td>kaskite</td>
<td>‘dig’</td>
</tr>
<tr>
<td>/teketi/</td>
<td>teksiu</td>
<td>tekite</td>
<td>‘flow’</td>
</tr>
<tr>
<td>/dreksti/</td>
<td>dreksiu</td>
<td>drekskite</td>
<td>‘bind’</td>
</tr>
</tbody>
</table>

The alternations in (4) reflect two operations: Metathesis, which exchanges a fricative and a velar stop before another consonant, and Degemination. Metathesis bleeds the application of Degemination in the imperative plural of ‘bind’, but feeds it in the 1st person singular future form. Derivations for ‘to bind’, ‘flow (IMP.PL)’ and ‘bind (IMP.PL)’ are shown in (5).

(5)  
<table>
<thead>
<tr>
<th>UR</th>
<th>/dresk+ti/</th>
<th>/tek+kite/</th>
<th>/dresk+kite/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metathesis</td>
<td>dreksti</td>
<td>n/a</td>
<td>drekskite</td>
</tr>
<tr>
<td>Degemination</td>
<td>n/a</td>
<td>tekite</td>
<td>n/a (bled)</td>
</tr>
<tr>
<td>SR</td>
<td>[dreksti]</td>
<td>[tekite]</td>
<td>[drekskite]</td>
</tr>
</tbody>
</table>

Turning to counterfeeding, it is useful following McCarthy (1999) to distinguish counterfeeding on the environment from counterfeeding on the focus. Counterfeeding on the environment may be illustrated by the interaction of Post-sonorant Voicing and Syncope in Tangale, a Chadic language spoken in Nigeria described by Kidda (1985, 1993), and Kenstowicz and Kidda (1987). Examples are given in (6). Vowels with an underdot are [RTR]. Suffixal vowels harmonize with root vowels
in this feature.

<table>
<thead>
<tr>
<th></th>
<th>UR</th>
<th>SG</th>
<th>DEF.SG</th>
<th>1SG.POSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/lọọ/</td>
<td>lọọ</td>
<td>lọọ-ị́</td>
<td>lọọ-nọ́</td>
</tr>
<tr>
<td></td>
<td>/ɓụɡat/</td>
<td>ɓụɡat</td>
<td>ɓụɡat-ị́</td>
<td>ɓuɡad-nọ́</td>
</tr>
<tr>
<td></td>
<td>/adụk/</td>
<td>adụk</td>
<td>adụk-ị́</td>
<td>aduɡ-nọ́</td>
</tr>
<tr>
<td></td>
<td>/wudó/</td>
<td>wudó</td>
<td>wud-ị́</td>
<td>wud-nọ́</td>
</tr>
<tr>
<td></td>
<td>/tụụžẹ/</td>
<td>tụụžẹ</td>
<td>tụụž-ị́</td>
<td>tụụž-nọ́</td>
</tr>
<tr>
<td></td>
<td>/lútu/</td>
<td>lútu</td>
<td>lút-ị́</td>
<td>lút-nọ́, *lúd-nọ́</td>
</tr>
<tr>
<td></td>
<td>/ɗụka/</td>
<td>ɗụka</td>
<td>ɗụk-ị́</td>
<td>ɗuɡ-nọ́</td>
</tr>
</tbody>
</table>

Preceding a sonorant consonant, stops surface as voiced. This is seen in the first person singular possessive forms for ‘window’ and ‘load’. A stem-final short vowel is elided when any phonological material follows (Kenstowicz, 1994, p. 96), as shown in all suffixed forms of ‘tooth’ and ‘horse’. Elision fails to feed Pre-sonorant Voicing. A derivation is given in (7).

<table>
<thead>
<tr>
<th></th>
<th>UR</th>
<th>/ɓụɡat+nọ́/</th>
<th>/tụụžẹ+nọ́/</th>
<th>/lútu+nọ́/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-sonorant Voicing</td>
<td>ɓụɡadnọ́</td>
<td>n/a</td>
<td>n/a (counterfed)</td>
<td></td>
</tr>
<tr>
<td>Syncope</td>
<td>n/a</td>
<td>tụụžnọ́</td>
<td>lútńó</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>[ɓụɡadnọ́]</td>
<td>[tụụžnọ́]</td>
<td>[lútńó]</td>
<td></td>
</tr>
</tbody>
</table>

Counterfeeding on the focus refers to chain shift phenomena. An example is furnished by raising in Nzebi, a Bantu language of Gabon (Guthrie, 1968; Clements, 1991; Kirchner, 1996). Before certain suffixes containing /i/, a vowel is raised, but only to the next level up. In other words, raising does not feed further raising. In rule-based terms, raising from mid to high must precede raising from low to mid.

The classic example of counterbleeding comes from Canadian English (Joos, 1942; Chambers, 1973). According to the standard account, the diphthongs /ɑ̄y/ and /ɑ̄w/ are raised to [ʌy] and [ʌw] before voiceless obstruents. Between a vowel and an unstressed vowel, the coronal stops /t d/ are both realized as a flap [ɾ]. To derive forms such as writing, however, Raising must apply before Flapping.

<table>
<thead>
<tr>
<th></th>
<th>UR</th>
<th>wayp</th>
<th>wayɾɲ</th>
<th>wipe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/hɑ̄yk/</td>
<td>ʰɑ̄yk</td>
<td>ʰɑ̄ykɲ</td>
<td>hike</td>
</tr>
<tr>
<td></td>
<td>/wed/</td>
<td>wed</td>
<td>weɾɲ</td>
<td>wade</td>
</tr>
<tr>
<td></td>
<td>/ʃut/</td>
<td>ʃut</td>
<td>ʃuɾɪŋ</td>
<td>shoot</td>
</tr>
<tr>
<td></td>
<td>/rɑ̄yd/</td>
<td>rɑ̄yd</td>
<td>rɑ̄yɾɪŋ</td>
<td>ride</td>
</tr>
</tbody>
</table>
Example (10) gives derivations for write, wading and writing.

(10) | UR      | /rāyt/ | /wed + ɪŋ/ | /rāyt + ɪŋ/ |
    | Raising | rāyt   | n/a        | rāyɪŋ       |
    | Flapping| n/a    | ɪn Jones   | rāyɪŋ       |
    | SR      | [rāyt] | [ɪn Jones] | [rāyɪŋ]     |

Up to this point we have considered the four possible two-rule interactions. There are also crucial three-rule interactions. One of these is double counterbleeding. The best known example is Yawelmani Yokuts (Newman, 1944; McCarthy, 1999). This language has a process of Vowel Harmony that spreads lip rounding from the stem onto a suffixal vowel of the same vowel height, e.g. *dub-hun ‘lead by the hand -NONFUT’ (cf. xil-hin ‘tangle-NONFUT’; bok’-hin ‘find-NONFUT’). This is counterbled by a process that lowers all underlying long high vowels to their mid counterparts. Lowering is then itself counterbled by a rule that shortens vowels in a closed syllable, e.g. *sap-al ‘burn-DUBITATIVE’, but *sap-hin ‘burn-NONFUT’. The effect of these rules on an underlying form like */ʔuːṭ-hin/ gives [ʔoṭhun].

Other kinds of three-rule interactions combine feeding/bleeding with counterfeeding/counterbleeding. Two kinds of interaction that have played an important part in recent debate are known by special names: rule-sandwiching and Duke-of-York derivations. The reason for singling out these two will become clear when we look in more detail at OT.

Rule sandwiching (Levi, 2000; Bye, 2001) refers to three-rule interactions of the form \( P > Q > R \), in which (a) rules \( P \) and \( R \) introduce identical structural changes, and (b) \( P \) feeds/bleeds (i.e. interacts transparently with) \( Q \), but \( R \) counterfeeds/counterbleeds \( Q \). Here we will illustrate the phenomenon with data from Mohawk, a Northern Iroquoian language spoken in New York State, Ontario, and Québec. The examples are in (11) taken from Hagstrom (1997, 114ff.). In this language, epenthetic vowels (shown underlined) may or may not be visible to stress assignment. The metrical visibility of epenthetic vowels is a function of the phonotactic conditions which trigger the epenthesis.

(11) /wak-haratat-u-haty/ wakharatatuhátie ‘I go along lifting up’
/k-ataʔkerahkw-haʔ/ kataʔkerákwaʔ ‘I float’
/a-k-r-ʌʔ/ ʌ́k erʌʔ ‘I will put it into a container’
/waʔ-t-k-atat-nak-ʔ/ waʔkatát enak eʔ ‘I scratched myself’
/wak-nyak-s/ wakényaks ‘I get married’
/te-k-ahsutr-haʔ/ tekahsutérhaʔ ‘I splice it’

Canonical stress falls on the penultimate syllable, as shown in the forms meaning ‘I go along lifting it up’, and ‘I float’. Mohawk has two rules that epenthize \( e \), one that operates in triliteral consonant clusters, and another that breaks up biliteral consonant clusters. Epenthesis into biliteral clusters is invisible to stress. Epenthizing once thus shifts the stress one position to the left of its canonical position. Thus, in the form meaning ‘I will put it into a container’, the stress is displaced to the antepenultimate syllable since the epenthetic vowel surfaces in the penultimate position. In ‘I scratched myself’, biliteral epenthesis has occurred twice, resulting in a second stress shift, with stress surfacing in the pre-antepenultimate syllable. Epenthesis into a triliteral cluster, however, feeds into stress placement, and stress surfaces in its canonical position. This is shown by the last two forms, meaning ‘I get married’ and ‘I splice it’. Stress Assignment is thus apparently “sandwiched” between two distinct epenthesis rules. (12) shows how this difference is derived in terms of an ordered sequence of rules.
Duke-of-York derivations (Pullum, 1976) are three-rule interactions of the form \( P > Q > R \), where \( P \) creates some string \( S \) that feeds \( Q \), and \( R \) counterbleeds \( Q \) by destroying the very string \( D \) that \( P \) created. It can be illustrated by Canadian French (Poliquin, 2006), data from which are shown in (13).

(13)  

<table>
<thead>
<tr>
<th>UR</th>
<th>/pœti/</th>
<th>pœ.ti'</th>
<th>petit</th>
<th>‘small’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/rabru/</td>
<td>w.a.bru</td>
<td>rabroue</td>
<td>‘scold’</td>
</tr>
<tr>
<td></td>
<td>/apik/</td>
<td>a.pik</td>
<td>à pic</td>
<td>‘steep’</td>
</tr>
<tr>
<td></td>
<td>/ekrul/</td>
<td>e.ksrul</td>
<td>écroule</td>
<td>‘crumble’</td>
</tr>
<tr>
<td></td>
<td>/filip/</td>
<td>fi.lip</td>
<td>Philippe</td>
<td>name</td>
</tr>
<tr>
<td></td>
<td>/tymuʃ/</td>
<td>tˤy.mʊʃ</td>
<td>tue-mouche</td>
<td>‘fly paper’</td>
</tr>
<tr>
<td></td>
<td>/elɛv/</td>
<td>e.leːv</td>
<td>élève</td>
<td>‘pupil’</td>
</tr>
<tr>
<td></td>
<td>/elɔʒ/</td>
<td>e.lɔːʒ</td>
<td>éloge</td>
<td>‘praise’</td>
</tr>
<tr>
<td></td>
<td>/eɡliz/</td>
<td>e.ɡliːz</td>
<td>église</td>
<td>‘church’</td>
</tr>
<tr>
<td></td>
<td>/misiv/</td>
<td>mɪ.siːv</td>
<td>missive</td>
<td>‘letter’</td>
</tr>
<tr>
<td></td>
<td>/difyz/</td>
<td>dɪ.fyːz</td>
<td>diffuse</td>
<td>‘diffuse’</td>
</tr>
</tbody>
</table>

In word-final open syllables, high vowels are tense, but in word-final closed syllables, high vowels are obligatorily lax, as can be seen from comparing the forms for ‘small’ and ‘scold’ with those for ‘steep’ and ‘crumble’. According to Poliquin’s interpretation, Final Closed Syllable Laxing feeds an optional rule of Vowel Harmony, which additionally introduces lax vowels into non-final open syllables. This is illustrated by the forms for ‘Philippe’ and ‘fly paper’. Vowels are lengthened preceding coda voiced fricatives (‘pupil’, ‘praise’). Now, lengthening does not generally give a tense vowel, but lengthened high vowels are always [+ATR], as in ‘church’. This is the crucial Duke-of-York U-turn, since it gives rise to surface disharmonic forms like mɪ.siːv and dɪ.fyːz. This, Poliquin argues, points to an intermediate representation *mɪ.sipv where Final Closed Syllable Laxing has applied, followed by Harmony. Laxing is subsequently reversed by Pre-voiced fricative Tensing. Derivations are given in (14) for à pic, Philippe, élève, and missive.

(14)  

<table>
<thead>
<tr>
<th>UR</th>
<th>/apik/</th>
<th>/filip/</th>
<th>/elev/</th>
<th>/misiv/</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>a.pik</td>
<td>fi.lip</td>
<td>n/a</td>
<td>mɪ.sipv</td>
</tr>
<tr>
<td>Vowel Harmony</td>
<td>n/a</td>
<td>fi.lip</td>
<td>n/a</td>
<td>mɪ.sipv</td>
</tr>
<tr>
<td>Pre-voiced fricative Tensing</td>
<td>n/a</td>
<td>n/a</td>
<td>e.leːv</td>
<td>mɪ.sipv</td>
</tr>
<tr>
<td>SR</td>
<td>[a.pik]</td>
<td>[fi.lip]</td>
<td>[e.leːv]</td>
<td>[mɪ.sipv]</td>
</tr>
</tbody>
</table>

2.2 Depth of derivation and Abstractness

Ideally every morpheme has a unique underlying representation. Rules allow us to minimize allomorphy by positing abstract underlying representations. This ideal breaks down in cases of suppletion, where two allomorphs are phonologically too distant to be credibly related by rule to the same un-
deriving representation, as is the case with *go* and *went* in English.\(^8\)

A classic demonstration of the need for abstract underlying representations, illustrated in (15), comes from Schane’s 1974 discussion of Palauan, a Malayo-Polynesian language.\(^9\) Stress falls on the last vowel of the stem when there is a suffix (in both the conservative and innovative future participle forms), otherwise on the penultimate vowel (in the present middle form). Unstressed vowels are realized as schwa.

\[(15) \quad \begin{array}{cccc}
\text{PRES.MIDDLE} & \text{FUT.PART} & \text{FUT.PART} & \text{UR} \\
& \text{(conservative)} & \text{(innovative)} & \\
mádəŋəb & dəŋəbáll & dəŋóbal & /dəŋəb/ \quad \text{‘cover’} \\
mátəʔəb & təʔəbáll & təʔíbəl & /təʔíb/ \quad \text{‘pull out’}
\end{array} \]

The stem of the verb for ‘cover’ has three surface allomorphs, [dáŋəb], [dəŋób], and [dəŋəb]. Selecting one of these surface allomorphs as the basic variant and attempting to derive the others from it will run in to problems for reasons that become clear when we consider the paradigm for ‘pull out’, which has the surface allomorphs [téʔəb], [təʔíb], and [təʔəb]. Vowel quality in unstressed syllables is predictable: it is always schwa. There is no similar generalization we can make with respect to vowel quality in the stressed syllable, however. The second vowels in the stems for ‘cover’ and ‘pull out’ must be specified respectively in the lexicon as /o/ and /i/. Since each occurrence of schwa can be understood as the result of applying a rule of vowel reduction, we can conclude that there is an abstract underlying form, /daŋəb/, which can be found by undoing its effects. Such examples are far from uncommon.

The same logic of reversing processes to arrive at the underlying form invites the possibility of positing segments that are never realized on the surface. Kiparsky (1973 [1968], 14) dubs this **absolute neutralization**, to distinguish it from the less controversial **contextual neutralization** where an underlying contrast is neutralized in some environments and preserved elsewhere. Absolute neutralization is a problem for acquisition because the learner is unable to appeal to surface morphophonological alternations to reconstruct the underlying representation of the segment; their only recourse is to the rules. On the assumption that rules of absolute neutralization were impossible or difficult to learn, Kiparsky proposed a stipulation, the Alternation Condition, which specified that “Neutralization processes cannot apply to all occurrences of a morpheme”. This was later revised to read “Neutralization processes apply only to derived forms” (Kiparsky, 1976). Nevertheless, several researchers proposed analyses in which they argued that absolute neutralization could allow considerable simplification of the grammar. A well-supported example is Hyman (1970), who proposed that Nupe, a Kwa language of Central Nigeria, has a rule lowering underlying mid vowels /ɛ/ and /ɔ/ to /a/ in all environments. The evidence for this abstract solution comes in part from the distribution of palatalized and labialized consonants. With one exception, palatalized consonants are found exclusively preceding front unrounded vowels, and labialized consonants before back rounded vowels. This distribution is thus consistent with being determined by rule. Preceding a low vowel /a/, however, a consonant may be plain, palatalized or labialized. Examples are shown in (16).
Hyman accounts for the distribution of secondary articulation by allowing lowering of the low mid vowels /e o/ to /a/ to counterbleed allophonic palatalization/labialization preceding a non-low vowel. He also cites evidence in favour of this analysis from the adaptation of Yoruba loanwords in Nupe, e.g. Yoruba tɔrɛ > Nupe tɔrɛ ‘to give a gift’. See Kiparsky (1976) for an overview of other similar proposals.

### 2.3 Natural and unnatural rule orders

Early generative phonology had a preoccupation with naturalness. This was obviously true in the development of distinctive feature theory and, to varying degrees, the search for formal mechanisms, such as non-linear representational frameworks (such as Autosegmental Phonology, Feature Geometry and Metrical Phonology), that would bring greater elegance and simplicity to the statement of natural rules. However, this same interest also disclosed itself in attempts to discern whether certain ordering relationships between rules could be said to be more natural or ‘unmarked’ than others. Related to the problem of absolute neutralization, is an argument by Schane (1968), who considers the interaction of Vowel Nasalization and Nasal Deletion in French. Nasal Deletion counterbleeds Vowel Nasalization, as in (17), which shows the derivations for bon ‘good’ and bonté ‘goodness’.

<table>
<thead>
<tr>
<th>(17)</th>
<th>UR</th>
<th>/bɔn/</th>
<th>/bɔnte/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Nasalization</td>
<td>bɔn</td>
<td>bɔnte</td>
<td></td>
</tr>
<tr>
<td>Nasal Deletion</td>
<td>bɔ</td>
<td>bɔte</td>
<td></td>
</tr>
</tbody>
</table>

Applying the same rules in the reverse order would, in addition to giving us the wrong result, destroy all the inputs to Vowel Nasalization, something that Schane calls absolute bleeding. Schane argues that orderings that would give this result are universally impossible.

Kiparsky (1968, 200) proposes what we might call the maximal application hypothesis: “Rules tend to shift into the order which allows their fullest utilization in the grammar”. Recalling the definitions in (1), feeding and counterbleeding both maximize rule application since, in each case, both P and Q apply. See also Kenstowicz and Kisseberth (1977, 159–168) for extensive discussion of Kiparsky’s proposal.

Arguing against his earlier proposal, Kiparsky (1971) offers an alternative naturalness metric that has surface distributions as its point of departure rather than the rule interactions themselves. He proposes (p. 623) that “rules tend to be ordered so as to become maximally transparent”. The intuition here is that, since the condition for transparent rules is present in the surface representation, this makes them less complex for learners to acquire. When the conditions are absent, the rule is said to be opaque. The following definition in (18) is adapted from Kiparsky (1973b, 79). (For
A process $P$ of the form $A \rightarrow B/C \_D$ is opaque if there are surface forms with any of the following characteristics:

a. instances of $A$ in the environment $C\_D$

b. instances of $B$ derived by the process $P$ that occur in environments other than $C\_D$

Since McCarthy (1999) it has also been customary to talk about ‘transparent’ and ‘opaque’ interactions. Feeding and bleeding are commonly grouped together as transparent interactions, counterfeeding and counterbleeding as opaque interactions.^{10} OT has brought with it several additional terms for talking about opacity. McCarthy and Prince (1995, 1999) adapts the terms underapplication (Type 1 opacity; (18-a)) and overapplication (Type 2 opacity; (18-b)) to refer respectively to the patterns derived in rule-based approaches by counterfeeding and counterbleeding (generically, misapplication).^{11} As we shall see in the next section, however, there may be reasons for severing the connection between rule interaction type and application type.

The idea that opaque rules were dispreferred was subsequently developed by Kisseberth (1973a) and Kenstowicz and Kisseberth (1973). Kenstowicz and Kisseberth (1973, 1977) argue specifically that the bleeding order should be considered unmarked in interactions between certain kinds of rules. For example, cross-linguistically epenthesis overwhelmingly bleeds voicing assimilation. The reverse ordering is to date only known from one language, New Julfa Armenian (Vaux, 1998).

Recently, there has been a renewal of interest in the question of natural rule ordering. A survey by Parker (2009) of rule interactions in seven languages finds clear preferences for transparent interactions, with feeding and bleeding respectively making up 54.8% and 24.4% of observed tokens. The incidence of counterbleeding and counterfeeding, at 11.8% and 9.0% is considerably lower. Ettlinger (2009), however, who specifically looks at the interaction of vowel harmony with other rules, finds that counterfeeding is very common, while bleeding is very rare.

A possible alternative to extrinsic ordering debated in the 1970s was the global rule (Kenstowicz and Kisseberth, 1970; Kisseberth, 1973a,b; Dinnsen, 1974; Kenstowicz and Kisseberth, 1977), a.k.a. derivational constraint. In standard derivational phonology, rules apply in strictly Markovian fashion to the current representation with no possibility of looking back to previous states or forward to optimal future states. A global reinterpretation of the writer vs. rider contrast, for example, might reformulate the Raising rule so that it only applies before underlyingly voiceless consonants. The forward-looking type of global rule was motivated by the idea that rules sharing a functional teleology could enter into conspiracies (Kisseberth, 1970a,b). Ultimately, global rules were rejected as having unwarranted generative power (see e.g. Kiparsky, 1976), although many of the issues would return in the 1990s with the advent of constraint-based approaches to phonology. The two closely related approaches described by Goldsmith (1993) and Lakoff (1993) propose rules that are sensitive to underlying level of representation. This conception also made it into the earliest attempt to deal with the problem of opacity in OT (McCarthy, 1996), which also forms part of the basis of Diagonal Correspondence Theory (Ettlinger, 2009).

### 2.4 Relations between rule interaction type and application type

In the previous section we suggested that there was a straightforward relation between rule interaction type and application type: feeding and bleeding translate into normal application, counterfeeding into underapplication, and counterbleeding into overapplication. Recent papers by Baković (2007,
to appear), Anttila et al. (2008) and Kavitskaya and Staroverov (2009), however, argue in one way or another for breaking this connection. A typology of the attested combinations of interaction and application type is provided by Baković (to appear).

Kavitskaya and Staroverov (2009) illustrate a type of rule interaction they dub fed counterfeeding, where $P$ feeds $Q$ and $Q$ counterfeeds $P$. In Lardil (Pama-Nyungan, Queensland; e.g. Hale, 1973), the absolutive forms of nouns whose stem is longer than two moras, a final vowel is deleted. Apocope feeds a rule that deletes final non-coronal consonants, giving derivations such as the following /dibirdibi/ $\rightarrow$ [dibirdib] $\rightarrow$ [dibirdi] ‘rock cod’. In derivations like this one, Consonant deletion counterfeeds Apocope. The fact that Apocope only applies in the absolutive form, however, raises the question whether this interpretation of the facts is the correct one. An alternative analysis in which the deletion of the final vowel was due to a morphological subtraction process, as assumed for example by Horwood (2001) and Kurisu (2001), would have to be carefully ruled out.

Anttila et al. (2008) discuss several processes affecting consonant clusters in Singapore English and an interaction Baković terms surface-true counterfeeding. Sibilant clusters are broken up by epenthesis, e.g. /reɪz+z/ $\rightarrow$ [reɪzəz] raises (cf. [bægz] bags), and the second of two obstruents in the coda is deleted, e.g. /test/ $\rightarrow$ [tes] test (cf. [testɪŋ] testing). Deletion thus creates potential inputs to Epenthesis, e.g. /list+z/ $\rightarrow$ [lisz], but instead of feeding Epenthesis, giving counterfactual *[lisəz], we find that Deletion feeds a rule of Degemination: /list+z/ $\rightarrow$ [lisz] $\rightarrow$ [lis]. As Baković (to appear) points out, Epenthesis is countered by Deletion, but it does not underapply since Degemination eliminates the input for Epenthesis, with the result that Epenthesis remains surface-true.

Bakovic further argues that certain types of feeding may result in opacity. Some alternations suggest self-destructive feeding, where $P$ feeds $Q$ and $Q$ counterbleeds $P$. The example Baković provides is from Turkish, where a putative rule that deletes [s] following a consonant feeds (and is counterbled by) a second rule eliding a [k] between vowels, e.g. /ajak+sɯ/ $\rightarrow$ [ajakɯ] $\rightarrow$ [aja.ɯ] ‘his foot’ (cf. [oda-sɯ] ‘his room’, /tʃan+sɯ/ $\rightarrow$ [tʃanɯ] ‘his bell’, /bebek+i/ $\rightarrow$ [bebe.i] ‘baby-ACC’). The alternation between /s/ and $∅$ is limited to the genitive and third person possessive suffixes, so it appears we are dealing with two allomorphs {-sl} (after vowels) and {-I} (after consonants) rather than a phonological rule. What is nevertheless interesting is that {-I} is still selected by consonant-final stems that lose their final consonant (opaque allomorph selection). Other s-initial suffixes do not evince the alternation, e.g. [fajda-sɯz] ‘without use’, [akɯl-sɯz] ‘without intelligence, stupid’.

Another type of opaque feeding comes from Classical Arabic, where words beginning with complex onsets prothesize a vowel. Since vowel-initial words are disallowed in Arabic, this further feeds the insertion of a glottal stop, e.g. /ktub/ $\rightarrow$ [uktub] $\rightarrow$ [ʔuktub] ‘write! (m.sg)’. This type of interaction, which Baković calls non-gratuitous feeding, entails a kind of overapplication: glottal epenthesis separates the epenthetic vowel from the word boundary that was part of the context of the rule, hence prothesis overapplies.

A case of mutual bleeding, or non-opaque counterbleeding, is the interaction of Epenthesis and Elision in Lardil. Clusters of i/u are broken up by insertion of a glide [w], e.g. /papi+w/ $\rightarrow$ [papiwɯ] ‘father’s mother -ACC.FUT’. In general, however, vowel clusters undergo simplification through deletion of the second vowel, e.g. /bæmpæ+w/ $\rightarrow$ [bæmpæ] ‘mother’s father -ACC.FUT’. Epenthesis bleeds Elision because it eliminates the vowel hiatus. It is also the case that Elision counterbleeds Epenthesis, but Epenthesis nevertheless cannot be said to overapply since Elision does not actually destroy the context for Epenthesis.

Baković’s most controversial proposal is cross-derivational feeding. Considerations of space preclude going into detail here, although see endnote 13.
2.5 No rule ordering and persistent rules

The preceding sections have looked at determinate sequential ordering, where rules apply in a strict sequence. In this section, we look at theories where the grammar does not specify an ordering. Koutsoudas (1976) distinguishes several possible conventions for rule application where no ordering is specified. Here we will concentrate on two, random sequential ordering and simultaneous application. These are still of relevance today also because of their similarities to serial and parallel implementations of OT.

In random sequential ordering, a rule or process applies whenever its structural description arises in the course of a derivation. The same structural description may arise more than once, and in this case the rule applies each time its input occurs. A rule with this property is known as persistent (Chafe, 1968). As already pointed out, rule persistence was a feature of the theory of rule interaction advanced by Koutsoudas et al. (1974 [1971]). It was also incorporated into Natural Generative Phonology (Vennemann, 1972; Hooper, 1976). In a grammar allowing only random sequential ordering, rules may only interact by feeding since, if $P$ creates an input to $Q$, then $Q$ must apply — the grammar does not have to specify the order $P > Q$. In cases of putative bleeding, however, the structural descriptions of both $P$ and $Q$ are simultaneously present in the input representation, requiring an arbitrary stipulation that $P$ applies first. In this connection Hooper (1976) discusses the interaction of epenthesis and voicing assimilation in English. In brief, the English plural suffix {-z} assimilates in voicing to a preceding voiceless obstruent (/kæt + z/ → [kaet]). When the stem ends in a sibilant, a vowel is inserted (/meiz + z/ → [meiziz]). But when the stem-final sibilant is voiceless, it is the epenthesis rule that has priority (/meis + z/ → [meisiz], *[meissis]). At best, random sequential application predicts that we should find variation between forms that are the outcome of bleeding interaction ([meisiz]) and those that result from counterbleeding ([meissis]). Hooper (p. 73f.) is therefore forced to adopt a solution in which the plural suffix is lexically specified as a disjunction of two allomorphs, {-iz} after sibilants, and {-z} elsewhere. If it was merely a matter of describing the variation in shape of a particular morpheme, this allomorphic solution would not be particularly costly. Unfortunately, it ends up missing several generalizations. For one thing, epenthesis in English is a general process that applies whenever the suffix is sufficiently similar to the stem-final consonant. The third person singular suffix thus behaves in exactly the same way as the plural, and the past tense suffix {-d} behaves analogously following a coronal stop (/paet + d/ → [paetd]; /paed + d/ → [paedid], *[paetd]). Without recourse to bleeding we would be forced to duplicate the allomorphic solution for each suffix. Second, as observed by Kenstowicz and Kisseberth (1977), epenthesis overwhelmingly bleeds voicing assimilation cross-linguistically, which would force us to duplicate the allomorphic solution across languages. Finally, the allomorphic solution could not deal with alternations in stems, since these pertain to open lexical classes, such as noun, adjective and verb.

We observed above that Metathesis may bleed Degemination in Lithuanian. Positing listed allomorphs for each stem that participates in the alternation (e.g. {dresk-} → {dreks-} ‘bind’) would not help deal with previously unencountered stems ending in $s \sim k$. Government Phonology (Kaye, 1992, 1995) also takes the line that all rules are persistent (Kaye, 1992, p. 141), and so similarly rules out the possibility of bleeding. However, this approach places great emphasis on representations in explaining phonological phenomena, which takes much of the descriptive brunt off rule interaction. Epenthesis, for example, is not dealt with in terms of processes, but by conditions on the phonetic interpretation of phonological structures, in which case the issue of ordering with respect to epenthesis does not arise (see Section 3 for an example).
Counterfeeding and counterbleeding are also incompatible with the logic of random sequential application that rules apply whenever their structural description is met. In counterfeeding, a rule $P$ fails to feed another rule $Q$ whose structural description it creates. Counterbleeding is problematic for the same reason as bleeding. As we shall see in Section 3, there has been much research in OT and Government Phonology to try and reinterpret opaque effects representationally or as non-phonological.\textsuperscript{14}

The basic premise of simultaneous rule application is that the underlying representation is the sole determinant of rule function. This entails that the application of a rule is not affected by the application of any other rule — there is no rule interaction. As Kenstowicz and Kisseberth (1979, 291–307) show, this entails that simultaneous rule application can only deal with counterbleeding and counterfeeding interactions, but not feeding or bleeding interactions.

Many of the issues regarding simultaneous and random sequential application we have been discussing return in a different guise in Optimality Theory (OT; Prince and Smolensky, 2004 [1993]; McCarthy and Prince, 1993). We will not go into details about how OT works here. For an introduction, see Uffmann (this volume). Here we will simply note some core properties of Classical parallelist OT that pose special problems for dealing with certain types of interaction.

First, in Classical OT, all processes apply simultaneously (in parallel) to the input form. However, this does not imply that OT and simultaneous rule application fare similarly with respect to capturing the two-rule interactions listed in (1) since simultaneous application is input-oriented, while OT is output-oriented. For this reason, simultaneous rule application can capture opacity effects, but OT cannot because the Faithfulness violation that results from the opacified process is not motivated by the surface structure. As a consequence of gratuitous violation of some Faithfulness constraint, a candidate representation $cand_{op}$ corresponding to the output of an opaque interaction will always be harmonically bounded by a more optimal transparent candidate $cand_{trans}$ with a proper subset of $cand_{op}$’s violation marks. In simultaneous rule application, on the other hand, the condition for the application of the process of course lies in the input to the rule. Solutions to the opacity problem in OT generally frame the issue as a ranking paradox. The desired winner $cand_{op}$ can be selected if there is some highly ranked constraint that $cand_{op}$ satisfies and its transparent competitor violates.

More recently, there has been increasing interest in a version of OT now known as Harmonic Serialism (HS-OT). HS-OT was first aired in Prince and Smolensky (2004 [1993], 94–95), but serious interest in exploring the possibilities of this implementation is very recent (McCarthy, 2008a,b). In HS-OT, processes apply in a random sequential order.\textsuperscript{15} GEN is limited to making a single harmonic improvement at a time (gradualness). Leaving aside for now the somewhat difficult problem of what should count as a single change, EVAL then determines the locally optimal output from this small set of candidates on the current pass, and the output is fed back to GEN, which produces another set of candidates for evaluation. This process continues to loop until the latest output of EVAL is identical to the latest input to GEN — if no further harmonically improving single edits can be made, the derivation is said to ‘converge’. HS-OT’s serialism is of little help from the point of view of solving the opacity problem since processes apply in a random order. At the very least what is required is some formal mechanism which would allow the OT grammar to mimic the extrinsic ordering of processes. Regardless whether the implementation is parallel or serial, developing the necessary mechanism is something of a technical challenge, although it has been attempted with some success for both parallel OT (Sympathy Theory; McCarthy, 1999) and HS-OT (Candidate Chains Theory, or OT-CC; McCarthy, 2006). Sympathy and OT-CC have been shown to be able to describe two-rule interactions, as well as multiple opacity as in Yawelmani Yokuts. But for quite fundamental reasons,
these attempts may represent the horizon of what opaque interactions it is possible to describe in OT. Specifically, even if we allow the linear ordering of processes, there would seem to be no way to allow the OT grammar to mimic linear ordering of rules, which would require a pairing of process and context. Recall that processes in GEN are context-free. Reintegrating processes with the contexts in which they apply would amount to abandoning OT.\textsuperscript{16} The implications of this are particularly apparent in the case of rule sandwiching, since what is at issue here is precisely getting the same process to apply at different points depending on context. Recall that in Mohawk, epenthesis of $e$ must apply crucially earlier in triliteral clusters than in biliteral clusters.

Another strategy for OT in dealing with opacity is to identify some function that opacity serves, allowing a solution to the constraint ranking paradox. Kaye (1974) argued that opacity should be considered natural, noting that it permits recovery of underlying contrasts. For example, the presence of the raised allophone in Canadian English [rayrɔː] signals that the flap is ‘really’ a voiceless obstruent, thereby providing an important clue for identification of the word. The idea that opacity may be functional in this way, as opposed to a mere epiphenomenon of sequential rule application, inspires pretty much every attempt to deal with opacity from the perspective of parallel OT (McCarthy, 1999, 2006; Goldrick, 2000; Łubowicz, 2003; Ettlinger, 2009). This also inherently limits the kinds of opaque interactions that recoverability-based approaches to opacity can deal with. Specifically, opaque contexts that are not present in underlying representation but arise through the application of processes, such as occurs in Duke-of-York derivations, cannot be dealt with (Idsardi, 1997). This is because the recovery of non-underlying opaque contexts would not contribute to recovering the underlying form. In sum, abstracting away from the details of any one implementation of OT, rule sandwiching and Duke-of-York derivations pose problems to both parallel and harmonic serial OT for reasons which have to do with the fundamental character of the theory. For this reason, the reality of both is a hotly contested area of contemporary debate over the relative merits of OT and rule-based phonology (e.g., Vaux, 2008). McCarthy (1999, 2006) argues that genuine Duke-of-York effects do not exist, while other researchers, including Idsardi (2000), Poliquin (2006), and Odden (2008), have argued that they represent real phenomena.

Let us see how each of the two-rule interactions described in section 2.1 fare in HS-OT.\textsuperscript{17} A corresponding OT analysis of the Norwegian feeding case presented in (2) might be constructed using the constraints in (19).

(19) \begin{tabular}{l}
\textbf{a. *CCC} \\
A cluster of three consonants is disallowed. \\
\textbf{b. MAX-C} \\
A consonant in the input must have a correspondent in the output. \\
\textbf{c. *r$\sim$Cor} \\
A sequence of /r/ followed by a coronal consonant is disallowed. \\
\textbf{d. UNIFORMITY} \\
A segment in the output must not have multiple correspondents in the input. \\
\end{tabular}

*CCC must outrank MAX-C and *r$\sim$Cor outrank UNIFORMITY. Even though the Markedness constraints *CCC and *r$\sim$Cor do not conflict, the logic of Harmonic Serialism additionally requires that the former outrank the latter. If they were unranked, the mapping from /skærpt/ to intermediate [skɔːrt] would not constitute a harmonic improvement: simplifying the cluster would merely trade one violation mark for another that the grammar deemed equally bad.

A serial evaluation is shown in (20). On the first pass, the locally optimal candidate satisfies
*CCC at the expense of incurring a lower-ranked violation of MAX-C, but it still violates lower ranked *r¬Cor. On the second pass, the locally optimal candidate from the first pass is bested by candidate (b), which additionally satisfies *r¬Cor, gaining a mark on lower-ranked UNIFORMITY.

(20) a. **First Pass**

<table>
<thead>
<tr>
<th>/skɑɾ₁p₂ + t₃/</th>
<th>*CCC</th>
<th>*r¬Cor</th>
<th>MAX-C</th>
<th>UNIFORMITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. skɑɾ₁p₂t₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. r skɑɾ₁t₃</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. **Second Pass**

<table>
<thead>
<tr>
<th>/skɑɾ₁p₂ + t₃/</th>
<th>*CCC</th>
<th>*r¬Cor</th>
<th>MAX-C</th>
<th>UNIFORMITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. skɑɾ₁t₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. r skat₁,₃</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Turning to the bleeding interaction of Metathesis and Degemination in Lithuanian (4), at least the following additional constraints in (21) would be needed.

(21) a. *s¬k¬C

A cluster of s¬k followed by a consonant is disallowed.

b. *GEMINATE

Geminates are disallowed.

c. **LINEARITY**

The output is consistent with the precedence structure of the input.

The tableaux in (22) illustrate the serial evaluation. The Markedness constraints *s¬k¬C and *GEMINATE both receive violation marks in the fully faithful candidate (a). Since the application of either Metathesis or Degemination eliminates both violation marks, the choice as to which process applies must be arbitrated entirely by the Faithfulness constraints LINEARITY and MAX-C. Ranking LINEARITY below MAX-C ensures that the candidate implementing Metathesis (c) wins over the candidate where Degemination has applied (b). On the second pass, no further harmonic improvements can be made, so the derivation converges.

(22) a. **First Pass**

<table>
<thead>
<tr>
<th>/dresk+kite/</th>
<th>*s¬k¬C</th>
<th>*GEMINATE</th>
<th>MAX-C</th>
<th>LINEARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. dreskkite</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. dreskite</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. r drekskite</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

b. **Second Pass**

<table>
<thead>
<tr>
<th>/dresk+kite/</th>
<th>*s¬k¬C</th>
<th>*GEMINATE</th>
<th>MAX-C</th>
<th>LINEARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. r drekskite</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Now let us address counterfeeding in Tangale (6). An OT analysis might use the constraints in (23).
The first constraint Elision is a shameless imitation of the corresponding rule; hardly a serious candidate for inclusion in the universal constraint set CON. It is nevertheless adopted here as an expository convenience.

(23) a. **Elision**  
A stem-final short vowel is disallowed preceding phonological material.

b. **Max-V**  
A vowel in the input must have a correspondent in the output.

c. **Pre-Sonorant Voicing**  
A sequence of voiceless obstruent followed by a sonorant is disallowed.

d. **IDENT[voice]**  
Corresponding segments must have identical specifications for [voice].

In order to derive the effects of Elision and Pre-sonorant voicing in isolation, Elision must outrank Max-V and Pre-Sonorant Voicing outrank IDENT[voice]. Next we have to get the two processes to interact. Simply combining these rankings, however, incorrectly predicts that Elision should feed Pre-sonorant Voicing. In Harmonic Serialism, processes apply persistently, until no further harmonically improving changes can be made. Assuming Elision dominates Post-SonVOI, the mapping from underlying /lútu+nó/ to the desired output form [lútnó] is harmonically improving. Unfortunately, the derivation does not converge at this point, because further harmonic improvements are possible. A second pass through GEN and EVAL converges on *[lúdnó], which does even better than the previous output and desired winner by transparently applying Pre-sonorant Voicing. In the serial evaluation shown in (24), the desired winner is indicated with ‘+’, the undesired but actual winner with ‘/x41’. The crucial fatal mark is indicated with an inverted exclamation point ‘¡’.

(24) a. **First Pass**

<table>
<thead>
<tr>
<th></th>
<th>/lútu+nó/</th>
<th>Elision</th>
<th>POST-SONVOI</th>
<th>MAX-V</th>
<th>IDENT[voice]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>lútunó</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>[lútnó]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. **Second Pass**

<table>
<thead>
<tr>
<th></th>
<th>/lútu+nó/</th>
<th>Elision</th>
<th>POST-SONVOI</th>
<th>MAX-V</th>
<th>IDENT[voice]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[lútnó]</td>
<td>j*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>[lúdnó]</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

McCarthy (2006) proposes an implementation of OT that essentially reintroduces extrinsic ordering of processes. The formal implementation of this idea is somewhat complex, so I shall limit myself to explaining the basic idea. McCarthy reifies the derivation as a kind of representation he dubs the **candidate chain**, which is subject to harmonic evaluation by **Eval**. The candidate chain is a linear ordering of forms, beginning with the fully faithful form and ending with the intended output. The chain may also contain intermediate representations. As in Harmonic Serialism, each form in the chain must represent a single harmonic improvement over the previous one. Markedness and Faithfulness constraints only evaluate the last form in the candidate chain. But by reifying the derivation in this way, it is possible to formulate constraints on relations between forms in the chain. In the Tangale
case, the comparison of two chains provides the focus of interest, the undesired *(lútnó, lútnó, lútnó), and the desired *(lútnó, lútnó). The chains differ in the way that violations of Faithfulness constraints are introduced, i.e., in the way they mimic the extrinsic ordering of processes. In both chains there is a violation of MAX-V (the result of a process of vowel deletion) in the second form of the chain, [lútnó]. In the first chain, however, there is also a violation of IDENT[voice] (the result of a process of voicing). McCarthy proposes that the first candidate chain can be ruled out by a constraint (one of a family called PRECEDENCE) that says violations of MAX-V cannot be followed in the chain by violations of IDENT[voice]. Crucially, this constraint is violated by the transparent chain, but not by the opaque chain. When included, the desired opaque winner is no longer harmonically bounded by the transparent candidate. Ranked at the appropriate position in the hierarchy, the newly identified constraint permits selection of the opaque candidate as the winner.

Finally, let us see how HS-OT fares on counterbleeding opacity. An OT analysis of the interaction of Canadian Raising and Flapping from (9) may be constructed using the following constraints.

\[(25) \]
a. RAISING
   The diphthongs /āy/ and /āw/ are disallowed preceding a voiceless obstruent.
b. IDENT[low]
   Corresponding segments must have identical specifications for [low].
c. *VTV
   A coronal stop is disallowed between vowels.
d. IDENT[sonorant]
   Corresponding segments must have identical specifications for [sonorant].

To derive the effects of Raising and Flapping in isolation, RAISING must outrank IDENT[low] and *VTV outrank IDENT[sonorant]. Again, simply combining the rankings predicts that the transparent candidate (c) should win. The flapping of /t/ satisfies both RAISING and *VTV in one fell swoop. Since the desired winner rʌyɾɪŋ entails two changes, it cannot be generated on the first pass. On the second pass, though, the optimal candidate is identical to the input from the previous pass, so the derivation converges incorrectly on the transparent form *rʌyɾɪŋ.

\[(26) \]
a. First Pass

<table>
<thead>
<tr>
<th>/rāy + ɪŋ/</th>
<th>RAISING</th>
<th>*VTV</th>
<th>IDENT[low]</th>
<th>IDENT[son]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rāytŋ</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. rāytŋ</td>
<td></td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. rāyɾɪŋ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

b. Second Pass

<table>
<thead>
<tr>
<th>/rāy + ɪŋ/</th>
<th>RAISING</th>
<th>*VTV</th>
<th>IDENT[low]</th>
<th>IDENT[son]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rāyɾɪŋ</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

The next section returns to transient rules and considers intrinsic ordering principles.
2.6 Intrinsic rule ordering

The 1970s saw a number of other significant attempts to restrict the power of extrinsic ordering. Intrinsic rule ordering refers to any universal principle that determines the sequential order in which rules apply, usually based on the formulation of the rules or the domain of their application (which is the topic of the next section). One of the earliest attempts at dispensing with extrinsic rule ordering was Koutsoudas, Sanders, and Noll (1974 [1971]). Their Proper Inclusion Precedence was eventually rechristened as the Elsewhere Condition (Kiparsky, 1973a, 1982, 1984), given in (27) according to Kiparsky’s formulation. The Elsewhere Condition assigns priority to disjunctively ordered pairs of rules in the same cycle (see below).\(^{18}\)

\[(27) \quad \text{The Elsewhere Condition (Kiparsky, 1984, p. 3)}\]

Rules A, B in the same component apply disjunctively if and only if

\begin{enumerate}
  \item The input of A is a proper subset of the input of B.
  \item The outputs of A and B are distinct. In that case, A (the particular rule) is applied first, and if it takes effect, then B (the general rule) is not applied.
\end{enumerate}

The Elsewhere Condition can be illustrated with data from Standard Finnish (cf. Kiparsky, 1973a, pp. 95f.). Finnish has a defective segment, here transcribed ‘/’, which is realized identically to a following consonant if there is one, and is deleted elsewhere (as in ‘go down’ and ‘go’). Phonologically, ‘/’ may be analyzed as a bare consonantal root node.\(^{19}\)

\[(28) \quad \begin{array}{lll}
  /\text{mene’ pois}/ & \text{menep pois} & \text{‘go away’} \\
  /\text{mene’ talolle}/ & \text{menet talolle} & \text{‘go to the house’} \\
  /\text{mene’ kotiin}/ & \text{menek kotiin} & \text{‘go home’} \\
  /\text{mene’ wasemmalle}/ & \text{menev wasemmalle} & \text{‘go to the left’} \\
  /\text{mene’ alas}/ & \text{mene alas} & \text{‘go down’} \\
  /\text{mene’}/ & \text{mene} & \text{‘go’} \\
\end{array}\]

Scobbie (1993) notes that ‘elsewhere’ in the general rule may trivially be reinterpreted as the negation of the context of the particular rule, allowing such interactions to be remodeled declaratively, i.e. without interaction. This is obviously impossible with rules whose contexts overlap, which requires extrinsic ordering.

2.7 Cyclicity

In theories of extrinsic ordering, each rule has only one chance to apply. In SPE, certain rules escape this restriction. The application of such cyclic rules is interleaved with processes that build successively larger morphological and syntactic constituents. Cyclic rule application begins with the innermost domain and proceeds outwards to the next most inclusive domain, and so on to the outermost. Certain nodes in morphological and syntactic structure, or the affixes that derive them, may be designated as cyclic. They trigger (re)application of cyclic rules (Brame, 1974). In English, for example, there is a basic division between suffixes that trigger reassignment of stress, and those that do not. Latinate suffixes typically trigger a new cycle, giving alternations like \textit{voluminous}~\textit{voluminous}. Germanic suffixes, on the other hand, generally do not work this way, cf. \textit{Hémulen}~\textit{Hémulénish}~\textit{Hémulenishness}.

Elision in Tangale is another example of a cyclic rule. Elision may create consonant clusters that
are repaired by epenthesis (Kidda, 1993, p. 44). Representative examples are given in (29). The epenthetic vowel is a high back rounded vowel [u] or [ụ] that harmonizes with the root for the feature [ATR].

(29) /tagza+nó/ taguznó ‘my sorghum’
    /lpré+źí/ lipurźi ‘your (f) needle’
    /sogló+gó/ sogulgó ‘your (m) fish’

The derivation of taguznó ‘my sorghum’ is given in (30).

(30) | UR | /tagza+nó/ |
    | Elision | tagznó |
    | Epenthesis | taguznó |
    | SR | taguznó |

Evidence for the cyclic application of Elision comes from verbs, where the same personal suffixes that mark possession on nouns are used to mark the direct object. Direct object markers may be followed by additional suffixes marking tense, and which trigger new applications of Elision. The data in (31a) from Kidda (1993, 42, 44) illustrate cyclic Elision in verbs followed by an object and tense clitic, and (31b) verb followed by tense clitic and noun object.

(31) a. /dóbé+nó+gó/ dóbúngó ‘called me’
    /sugdé+źí+gó/ sugdúžgó ‘pierced you (f)’
    /káyé+gó/ káyuğgó ‘chased you (m)’

b. /poné+gó/ pongó /poné+gó#ńaj/ ponuñ ńaj ‘knew (Ngai)’
    /áné+gó/ angó /áné+gó#ńaj/ anuñ ńaj ‘paid (Ngai)’
    /timlé+gó/ timulgó /timlé+gó#ńaj/ timlug ńaj ‘hypnotized (Ngai)’

The correct distribution of epenthetic vowels can be derived by assuming that Epenthesisis not a cyclic rule, applying instead after all the cyclic rules. The resulting derivation is shown in (32). Double brackets [ ] show the morphological bracketing. As the derivation moves outwards, the brackets of the previous cycle are erased.

(32) | UR | [ timlé go ] [ [ timlé go ] ńaj ] [ [ sugdé źí ] gó ] |
    | First Cycle |
    | Elision | timlgó | timlgó | sugdží |
    | Second Cycle |
    | Elision | n/a | timlg ńaj | sugdžgó |
    | Post-cyclic rules |
    | Epenthesis | timulgó | timlug ńaj | sugdúžgó |
    | SR | timulgó | timlug ńaj | sugdúžgó |

Allowing Epenthesisis to apply cyclically would predict the wrong result for ‘hypnotized Ngai’, as shown in (33).
Having to stipulate which rules are cyclic and which are not may seem like an undesirable enrichment of the theory. Because Elision specifically targets stem- or domain-final vowels preceding other phonological material, it is an example of a process that only applies in a morphologically derived environment (Kiparsky, 1976; Mascaró, 1976). Domain-finality does not have to be inscribed into the formulation of the rule. It falls out naturally from the assumption that it is a cyclic rule. Epenthesis does not have this restriction. Suppose we were to try and eliminate the distinction by claiming that all rules were cyclic. Then we might attempt to replace Elision with a purely phonological Syncope rule that targets any medial light syllables irrespective of domain. This would serve to get rid of the unwanted epenthetic vowel in the counterfactual output *[timulgŋaị] in (33), but it would also counterfactually predict the deletion of medial light vowels in morphologically non-derived environments. The existence of trisyllabic words such as dānjik ‘blouse’, lanóró ‘donkey’ and mbipíde ‘cobra’, all with a medial light syllable intact, proves this assumption wrong: the medial light syllable has to be the result of the concatenation of morphemes in order for it to count as a target for deletion. In other words, we’re back to our original Elision rule.

Cyclic Phonology was further elaborated in Lexical Phonology (Kiparsky, 1982, 1984, 1985; Kaisse and Shaw, 1985; Mohannan, 1986), which attempted to formulate the general properties of cyclic and other levels, also known as strata. The current consensus is that there are three strata, two lexical and one post-lexical. The lexical strata are the stem level (cyclic) and the word level (post-cyclic). The counterbleeding interaction of Canadian Raising and Flapping requires that Raising apply in an earlier stratum than Flapping. The ascription to stratum should follow from the morphosyntactic domain in which the process applies. Since Flapping applies across word-boundaries (as in *my mother bough[r] a parrot), it can be ascribed to the post-lexical component. Raising, which doesn’t look beyond the word for its context, is lexical. For our purposes, a more specific level ascription is unnecessary (although see Bermúdez-Otero (2004) for an argument that Raising should be assigned to the Stem Level). Another important innovation with respect to Cyclic Phonology was that Lexical Phonology also manifested a concern with the properties of levels as levels, as opposed to mere steps in a derivation. Lexical rules, for example, are restricted to outputting structures that are also possible in underlying representation: they must be structure preserving. Post-lexical rules may produce allophones, even gradient outputs. Lexical rules may have exceptions, but post-lexical rules apply without exception. Finally, post-lexical rules are blind to domain, applying also across word boundaries. Since Canadian Raising is lexical, it must be structure preserving, i.e. it outputs segments that are possible underlying forms. As we shall see in Section 3, this prediction is borne out by the existence of words like tiger, idiosyncratically with [ʌy], which lack the Raising context.

Stratal OT is an adaptation of the central ideas of Lexical Phonology in OT (Kiparsky, 2000, 2002; Bermúdez-Otero, to appear). Each stratum corresponds to an OT grammar, each a total ranking of the universal set of constraints. An important difference between rule-based Lexical Phonology
and Stratal OT, however, is the predictions each makes with respect to within-cycle phonological opacity. In rule-based LP, each stratum consists of an extrinsically ordered battery of rules, allowing the capture of within-cycle counterfeeding and counterbleeding relationships. In Stratal OT, however, each stratum is transparent. The tableaux in (34) illustrate Stratal OT applied to the interaction of Canadian Raising and Flapping. The strata differ in the relative rankings of IDENT[son] and *VTV. On the Stem level, Flapping does not take place because IDENT[son] outranks *VTV. This ranking is reversed on the Post-lexical level, which produces Flapping. The overapplication effect is derived by ranking IDENT[low] high on the Post-lexical level, which forces identity of vowel height with the output of the previous level.

(34) a. Stem Level

<table>
<thead>
<tr>
<th>/rɑ̄yt + iŋ/</th>
<th>RAISING</th>
<th>IDENT[son]</th>
<th>IDENT[low]</th>
<th>*VTV</th>
<th>*λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rɑ̄ytŋ</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. rɑ̄ytŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. rɑ̄ytŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. rɑ̄ytŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Post-lexical Level

<table>
<thead>
<tr>
<th>rɑ̄ytŋ</th>
<th>RAISING</th>
<th>*VTV</th>
<th>IDENT[low]</th>
<th>IDENT[son]</th>
<th>*λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rɑ̄ytŋ</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. rɑ̄ytŋ</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. rɑ̄ytŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d. rɑ̄ytŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Parallel OT has developed other approaches to cyclic effects, such as Transderivational Correspondence Theory (Benua, 1998). Actually occurring words, like write and writing, enter into correspondence relations with each other, permitting the formulation of Output-Output Faithfulness constraints that require identity (in some dimension) of bases and their derivatives.

3 Alternatives to opaque interaction

In the previous section we compared and contrasted theories for which the ordering of processes or rules has to specified in the grammar of a particular language (extrinsic ordering) and theories in which the sequence of application is made to follow from principles of Universal Grammar (intrinsic ordering). From the point of view of explaining language acquisition, theories where as much as possible derives from universal principles are to be preferred, for the reason that they impose restrictions on the space of hypotheses that the learner can posit in the course of constructing the grammar of their native language. The kind of theory that is most interesting from the point of view of explaining language acquisition, however, is not necessarily best from the perspective of attaining maximal empirical coverage. This tension between the demands of explanation and description are evident in the debates on extrinsic and intrinsic ordering. Extrinsic ordering is explanatorily costly
because it enlarges the learner’s space of possible solutions, but perhaps necessary, since nothing else can account for the attested range of process or rule interactions. Opponents of extrinsic ordering have appealed to four kinds of alternative account, listed in (35); which, if any, of these strategies is appropriate can only be determined on a case by case basis, however. Specific types of opacity do not have any single type of reinterpretation.

(35)  a. phonological reinterpretation (representations; other rules)
b. phonetics
c. allomorphy/suppletion
d. analogy/lexicon

Given an alternation A~B, and a hypothesized rule A→B/C D, one solution is to retain the alternation in the phonology but give it a different interpretation. We can identify two subtypes of this strategy. The first is to try and replace the opaque rule with a transparent one. Sometimes there may be an alternative surface-apparent condition that we have missed which allows us to write a different, fully transparent, rule, say, A→B/E F. The second is to reinterpret the alternation in representational terms so that, despite appearances, the ‘opaque’ structure does not satisfy the structural description of the rule. The three remaining strategies all involve reallocating the alternation to a different component. For example, it may not be the case that B should figure in the discrete surface representation at all, but be allowed to emerge from the way this representation is mapped onto continuous parametric representations in the phonetic component (e.g. Browman and Goldstein, 1986). The last two strategies may be traced to the work of Miłojak Kruszewski and the Kazan School (Kruszewski, 1881; Baudouin de Courtenay, 1972 [1895]; Klausenburger, 1978; Anderson, 1985). Closer investigation may reveal that B is not a phonological alternant of A (approximating Kruszewski’s ‘divergent’), but that A and B involve suppletive allomorphy (cf. Kruszewski’s ‘correlative’). We will exemplify each of these in turn using examples adduced earlier in the chapter.

Staying within the phonological ambit for the time being, consider a potential transparent reinterpretation of an opaque rule. Poliquin’s argument for Duke-of-York derivations in Canadian French depends on understanding laxing in non-final high vowels as the result of a Vowel Harmony rule.20 As interpreted by Poliquin (p. 179), a high vowel in an open syllable becomes [−ATR] when separated by zero or more syllables from a word-final closed syllable containing a [±high, −ATR] vowel. The rule formulated by Poliquin is given in (36).

(36)  [+high] → [−ATR]  

Interpreted this way, the rule of Pre-voiced Fricative Tensing inevitably destroys part of the context for the application of Vowel Harmony, and the result is apparent counterbleeding. Another interpretation of the facts, though, is that the required rule is not a Vowel Harmony rule at all, but an entirely transparent rule that changes a non-final high-vowel to [−ATR] preceding any high vowel in the last syllable, irrespective of its specification for [ATR].

The handling of vowel~zero alternations in Government Phonology may serve as an illustration of how representations may be used to reanalyze putative opacity effects. Consider once again the facts of Elision and Epenthesis in Tangale. Building on Nikiema (1989a,b), Charette (1991, 106–112) provides an analysis of Tangale Elision in which the vowel is delinked from the nucleus but the position itself remains — as required by the principle of Government Phonology that every representation begin with an O(nset) and end in a R(hyme). For dobono ‘call me’ and dobungo ‘called me’, Charette proposes the structures in (37). When the empty nucleus is properly governed by a following nucleus
with melodic content, it has no phonetic interpretation. When it is not properly governed, the empty nucleus is interpreted as [u].

(37)$$\begin{array}{ccc}
R & R & R & R & R & R \\
N & N & N & N & N & N \\
\times & \times & \times & \times & \times & \times \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
[ [ d o b e ] n o ] & [ [ d o b u ] n o ] & [ g o ]
\end{array}$$

Although the question is not addressed by Charette, or Nikiema, it is now clear how we might represent the difference between ɓụgadnọ́, where Post-sonorant Voicing applies, and lútnọ, where it does not. Leaving aside the question of which feature is involved in the assimilation (although, see Nasukawa, 2005), it is possible to say that Post-Sonorant Voicing fails to apply because the intervening empty nucleus inhibits spreading of the relevant feature. Solutions of this kind, then, redefine the representation in such a way that the structural description of the would-be opaque rule cannot be met.

(38)$$\begin{array}{ccc}
R & R & R & R & R \\
N & N & N & N & N \\
\times & \times & \times & \times & \times \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
ɓ & u & g & a & d & n & o & l & u & t & u & n & o
\end{array}$$

McCarthy’s (2006) response to putative rule sandwiching in Mohawk illustrates how certain phenomena might be reassigned to phonetic interpretation. The way in which the phasing of articulatory gestures may give rise to apparent deletions and insertions is a fruitful area of research activity (Brownman and Goldstein, 1986). Overlapping gestures may partially or completely mask certain segments, even though represented by the appropriate gesture. In a phrase like ten pin, the labial closure gesture will in casual speech overlap with the preceding nasal, resulting in something that sounds like [tem pin], although /n/ retains an apical gesture in articulation. In such cases, it would be inappropriate to speak of the segment as having been deleted. Similarly, incomplete overlap, or open transitions between gestures may result in the appearance of excrecent vowels (cf. Levin, 1987). Such vowels are not epenthetic in a phonological sense, since they are not present in the surface representation. Hall (2003) argues that excrescence is more likely to occur in transitions to voiced sounds. McCarthy observes that what we called Biliteral Epenthesis above, the second consonant must be one of the resonants /nrw/. He argues that Mohawk has only one epenthesis rule, Triliteral Epenthesis, which interacts transparently with stress. The e that is invisible to stress is simply an excrecent vowel with no phonological representation.

This brings us to the morpholexical solution. Several attempts have been made from within the OT camp, including Sanders (2003), Mielke et al. (2003), and Green (2004, 2007). As we saw in Section 1, the opaque interaction of Canadian Raising and Flapping played a major role in the emergence of the generative paradigm in phonology. Attempts to eliminate the opacity date right back to Joos’ seminal paper (Joos, 1942). Kaye (2009) presents a recent argument against opaque Canadian Raising, which he argues is an artefact of the assumption that the diphthongs in ride and write are underlingly the
same. He notes first of all that, considered apart from any high off-glides, the vowels [ʌ] and [ɑ] are contrastive in Canadian English, as evidenced by minimal pairs such as [kʌˈb] cub vs. [kɑ̄ˈb] cob and [kɑ̄ˈt] cut vs. [kɔɪˈt] cot. Given the independent existence within the system of both of these vowels, Kaye suggests (p. 17) “[t]he obvious analysis is that C[anadian] E[nglish] “right” and “out” have rʌyt and ʌwt as their lexical representations”, that is, the initial vowels of [ˈʌy] and [ˈaw] now belong to the phoneme /ʌ/, not /ɑ̄/. Kaye concludes that deriving [ˈʌy] from [ˈoɪy] would introduce an unnecessary disparity between the underlying and surface levels: the qualitative difference therefore cannot be allophonic. If the traditional phonological account is correct, however, we would expect to find [ˈɔy] and [ˈaw] limited to the environment preceding a voiceless obstruent. What we find instead, at least for [ˈɔy], is that it can occur lexically in non-raising environments. Kaye reports that some speakers use this quality in tiger. K. Currie Hall (2005), working within a stochastic, exemplar-based approach, also reports a breakdown, especially noticeable in the case of polysyllabic words, in the predictability of the distribution of [ɔɪy] and [ˈʌy]. She demonstrates that lexical neighbours (words that share similar or identical substrings) may in fact exert an analogical influence on the choice of the diphthong in a way that runs counter to the putative phonological generalizations. Thus, gigantic and angina unexpectedly have [dʒʌɪ], despite the absence of a voiceless obstruent in the relevant syllable, and Siberia and psychology have [sʌɪ]. Completing Kaye’s point, she shows that [ɔɪy] turns up in syllables closed by a voiceless obstruent, such as like and life. In her 2005 paper, Currie Hall suggests factoring lexical neighbourhoods into the description of the phonological rule, but in her 2009 dissertation she essentially opts for relexicalization.

Perhaps having contrast in 4% of environments (before [r] in words like writing/riding) has opened the door for new generalizations to emerge. For example, language users could generalize that [ai] is possible before voiced segments and extend that to other words like gigantic. The prediction then is that [ɔi] and [ɔi] could continue along the continuum and end up being entirely unpredictably distributed: Fully contrastive.

(Conrie Hall, 2009, p. 124)

This perspective opens up interesting possibilities for understanding structure preservation phenomena. Recall from the previous discussion that opaque rules are necessarily lexical and, being lexical, they can only output contrastive segments of the language. When allophones are opacified, therefore, they are necessarily reanalyzed as phonemes. However, structure preservation is only a stipulated property of lexical rules. The ultimate reason for the breakdown observed by K. Currie Hall may be that children acquire their knowledge of phonological distributions before undertaking any morphological analysis of the input, essentially applying structuralist discovery procedures to construct the inventory of contrastive segments in their language. Some words thus come over time to be lexicalized with the ‘wrong’ allophone. One possible mechanism underlying errors of this type may be analogy with forms where the allophonic rule has overapplied. Alternatively, in the absence of a 100% reliable generalization learners may be unsure about which sound category to assign a given word, especially one that they do not encounter frequently. When the child later acquires the morphology and opaque rules along with it, it is too late to undo relexicalization. The possibility of relexicalization raises the interesting question whether children do in fact acquire opaque rules at all or whether the generalizations are simply part of the lexicon. See however McCarthy (2006) and Ettlinger (2009) for arguments for the psychological reality and learnability of opaque rules.

What is allophonic according to Kaye is the variation in the length of the vowel: all stress-bearing vowels in monosyllables lengthen in what Kaye calls the ‘non-fortis environment’, which is to say at
the end of a word, optionally preceding a consonant not drawn from the set of voiceless obstruents. Thus [bæt] \textit{bat} vs. [bæˑd] \textit{bad}, [kɑ̄t] \textit{cot} vs. [kɑ̄ˑd] \textit{cod}. This kind of variation is entirely general in English. Since Kaye’s Non-fortis Lengthening and Flapping are mutually non-affecting, however, their relative ordering is simply not crucial.\textsuperscript{22}

4 Conclusions

A derivation is a mapping between an underlying and a surface level, each of which is made up of the same basic type of representational elements. Derivations also entail the application of processes that potentially alter underlying structure in destructive ways. As the field has developed, the relative emphasis on levels and derivations has shifted. At certain times, most obviously at the inception of the generative programme in phonology, a strong emphasis on derivations, processes and rules has tended to divert attention away from an interest in the properties and functions of levels and their interrelation. Nevertheless, a concern with levels has repeatedly reasserted itself, and changes in the conception of levels have brought about corresponding changes in the understanding of derivations. The fundamental question remains whether linguistically significant generalizations should be stated as constraints on one or more levels, or as rules deriving one level from another. It is interesting in this connection to compare rule-based phonology and OT in their earliest and most radical versions with the more moderate positions most practitioners of each approach have come to adopt in the light of experience with real data. Most versions of rule-based phonology came to incorporate both restrictions on underlying representations (morpheme structure conditions), surface representations (output constraints such as the OCP), and outputs of lexical strata (Structure Preservation). OT has in its turn had to find ways of mimicking the extrinsic ordering of processes to deal with the opacity problem. Perhaps it is too optimistic to hope that a consensus on these fundamental issues will emerge any time soon. But between these two ways of understanding the grammar is a creative tension that will surely spur us on to novel insights about the nature of language for decades to come.

Notes

\textsuperscript{1}Several early implementations of generative grammar do in fact posit uninterpretable symbols in underlying or deep structure. Such symbols had to be deleted during the course of the derivation. Indeed, the derivation was partly driven by the requirement that such uninterpretable symbols be absent at the surface. In Chomsky (1965), for example, sentences that had not (yet) been incorporated into a higher matrix clause were bounded by the juncture symbol \#. Every occurrence of \# in the deep structure \textit{John saw the man} \# \textit{the man kicked the cat} \# must be eliminated by applying a relative clause transformation to derive \textit{John saw the man who kicked the cat}. Similarly, in Chapter 9 of SPE, distinctive features could only have values of + or − at the surface, but underlyingly they might also be specified as ‘marked’ \(m\) or ‘unmarked’ \(u\). In the course of the derivation, these values were replaced by + and − through what were known as marking conventions. The notion of unmarkedness continued to develop in Underspecification Theory (see Steriade, 1995, for an overview) and representational theories allowing only unary features, such as Government Phonology.

\textsuperscript{2}As Goldsmith notes, this conception had a couple of ready metaphors to hand, in the derivation
from proto-forms in historical linguistics and in production systems used in artificial intelligence.

3Other terms commonly in use in the literature for ‘focus’ and ‘environment’ include, for ‘focus’, **target**, and for ‘environment’, **context**, **determinant**, or **trigger**.

4The origins of this conception are due to Bloomfield (1939), who cites the *Aṣṭadhyāyi* of Pāṇini as his chief inspiration, and Wells (1949). See Goldsmith (2008) for discussion of pre-generative antecedents of this concept.

5These terms are ultimately due to Kiparsky (1968).

6McCarthy (2007a, 101) proposes the term **designated level** to refer to those representations “with special restrictions on their content or unique roles to play”. In addition to the underlying and surface levels, the output of a cycle also constitutes such a designated level. See Section 2.7 below.

7A vacuous Duke-of-York derivation would skip \( Q \), and \( R \) would destroy \( S \) as soon as \( P \) created it. While possible formally, the learner would lack the crucial evidence to posit such an interaction.

8As the Austronesian scholar Ken Rehg comments in a paper (Rehg, 2001, p. 218), though, “one of the imperatives of generative phonology — that allomorphy must be minimised — is sometimes at odds with the data, typically in very subtle ways”. Suppletion is easy to diagnose when the variants in question have little in common phonetically, but suppletion cannot be ruled out even in the case of phonetically similar variants. There is a widespread assumption that phonetically similar allomorphs must derive from the same underlying form, and this very assumption will tend to favour analyses with abstract underlying representations and rule interaction. Researchers working within Government Phonology have also pointed out that adhering to this assumption in all cases may run into empirical problems, not to mention the theoretical issues raised by abstractness and extrinsic ordering.

9Schane cites a dissertation by Flora (1974).

10As a terminological caveat, note that a pair of rules may interact transparently, but either of the rules may be opaque because it is opacified by some third rule.

11These terms are originally due to Wilbur 1973.

12A further logical type is what Koutsoudas (1976, 4) terms **arbitrary ordering**, which resembles random ordering except that the rules may not apply more than once.

13Baković (2007, to appear) argues that this does not represent a genuine case of bleeding. Bakovic raises concerns about implicit duplication in the traditional way of understanding the environment of epenthesis as applying between two consonants that potentially differ only in their value for the feature [voice]. He writes: ‘Epenthesis must arbitrarily ignore the difference in voicing between the stem-final /t/ and the suffix /d/ — precisely the difference that would be neutralized by Assimilation were it to apply. A more explanatorily adequate analysis would eliminate this redundancy, making **strict identity** a requirement on Epenthesis [...] and relying on Assimilation to provide the necessary context [...].’ Baković uses this interpretation of epenthesis to argue for a type of opacity he dubs **cross-derivational feeding**. Showing how this works would take us too far afield, but suffice it to mention that his concern is to show that OT is able to handle some types of opacity that rule based theories cannot. **Contra** Baković, I can’t find any reason for thinking that allowing epenthesis to ignore [voice] is arbitrary in any way. That homorganic similarity rather than complete segmental identity is the correct interpretation is also suggested by the fact that epenthesis in English applies
to break up clusters of sibilants that potentially differ not only in their specification for [voice] but also in their value for the feature [anterior], e.g. [bʊʃɪz] bushes. In the case of [anterior], the issue of duplication does not arise, since there is no phonological assimilation rule in English involving this feature.

14Myers (1991) argues for an approach with a division of labour between extrinsically ordered and persistent rules, stating (p. 339) ‘the set of persistent rules specifies the inventory of phonological elements (segments, syllables and feet) in a language, whereas the ordered rules specify the distribution of those elements’.

15McCarthy (2007b) originally proposed the name ‘Persistent OT’.

16A partial solution to this problem is offered by Ettlinger’s 2009 Diagonal Correspondence Theory (DCT). DCT builds on proposals in the literature that segments within the same harmonic span may be ‘syntagmatic correspondents’. In this way, the focus of the autosegmental spreading of some feature may be the ‘diagonal correspondent’ of that feature’s underlying determinant. The underlying determinant of a spreading process may thus have diagonal correspondents in the surface representation whether or not it has ‘perpendicular’ correspondents (it has been deleted), and whether or not its perpendicular correspondent has identical values for the harmonizing feature. This allows Ettlinger to describe opaque spreading rules. It is not clear how it may be extended to other types of rule, however.

17There will always be other plausible candidates that would need to be excluded by bringing on additional constraints in a fully worked out account. In the interests of space, though, we abstract away from this here and simply offer the barebones of an analysis for each case.

18Other intrinsic principles include Archangeli’s (1984, 51) Redundancy Rule Ordering Constraint, which forces all redundancy rules to apply as late in the derivation as possible.

19Kiparsky identifies the segment as /k/, but this choice appears to be synchronically arbitrary. The only evidence available to the child is that the segment is an unspecified consonant of some kind.

20I am extremely grateful to an anonymous reviewer for bringing this important point to my attention.

21The relexicalization argument is rendered plausible by the existence of other clear cases of contextual relexicalization in the history of English. In certain parts of the English-speaking world, including the southeast of England, southern New England and eastern Virginia, [ə] was lengthened to [æː] and subsequently retracted to [æː] before the fricatives [θʃs] and coronal nasal+obstruent clusters [nt nd ns nj]. Despite the importance of context, the sound change was clearly not allophonic, as evidenced by the existence of lexical items that failed to undergo reassignment, such as asp, mass, and pant. The replacement of /æ/ with /ɑː/ in many words is a clear case of lexical diffusion (Wang, 1968).

22Idsardi (to appear) nevertheless provides evidence in a recent account that Canadian Raising is a productive rule that results in actual alternations, e.g. [ai] i vs. [aiθ] i-th and [wai] y vs. [waiθ] y-th. He also mentions that his own variety displays raising before certain clitic prepositions, e.g. [laɪ] lie vs. [laɪɾə mi] lie to me.
References

Bloomfield, L., (1933), Language. Chicago: Chicago University Press.

30


Morning Calm. Hanshin, Seoul: Linguistic Society of Korea, pp. 3–91.
och Wiksell, pp. 135–164.
291–306.
and A. L. Vanek (eds.), Studies Presented to Robert B. Lees by his Students. Edmonton, Alberta:
Linguistics. Papers in Honor of Henry and Renée Kahane. Urbana, IL: University of Illinois Press,
pp. 418–441.
Kenstowicz and C. Kisseberth (eds.), Issues in Phonological Theory, [= Janua Linguarum, Series
major, 74]. The Hague: Mouton.
Historiographica Linguistica, 5, 109–120.
Koskenniemi, K., (1983), Two-level Morphology: A General Computational Model for Word-form
Recognition and Production. Technical report, University of Helsinki. Publication no. 11.
Language, 50, 1–28. Originally distributed as mimeo by Indiana University Linguistics Club,
Bloomington, IN.
Kurisu, K., (2001), The Phonology of Morpheme Realization. Doctoral Dissertation, University of
Phonology Conference 1, Montréal, April 2000.
Coast Conference of Formal Linguistics.


Newman, S., (1944), The Yokuts Language of California, [=The Viking Fund Publications in Anthropology, 2]. The Viking Fund.


Wilbur, R., (1973), The Phonology of Reduplication. Doctoral Dissertation, University of Illinois at Urbana-Champaign.