Derivational Residue: Hidden Rules in Optimality Theory

Darlene LaCharité and Carole Paradis

6.1 Introduction

Optimality Theory (OT) (following McCarthy and Prince 1993; Prince and Smolensky 1993) has focused on the identification and articulation of ranked phonological constraints by test-driving phonological phenomena from a variety of languages through selected subsets of those constraints. OT’s goal has been to account for phonological generalizations within languages, and for differences among languages, using universal surface constraints that are prioritized on a language-specific basis, instead of using rules, derivations and intermediate representations, as in Chomsky and Halle (1968) (henceforth SPE).

There is now general agreement that shifting the focus from rules to constraints has been remarkably successful. Perhaps because of that success, some of OT’s proponents have concluded that OT has replaced rules with ranked constraints entirely. For example, when comparing OT with phonotactic and repair theories, including Harmonic Phonology (following Goldsmith 1989; Goldsmith and Larson 1990; Wiltshire 1992; Goldsmith 1993; et al.) and the Theory of Constraints and Repair Strategies (henceforth TCRS) (following Paradis 1988a, 1988b, 1990; LaCharité and Paradis 1993; et al.), Prince and Smolensky (1993: 203), two of OT’s principal architects, imply that the results of rules are achieved in OT solely through the interaction of violable constraints. Itô et al. (1995: 578) say explicitly that in OT ‘there are no rules or repair strategies, and no serial derivation’. Yip (1993: 2) asserts that an OT account of onsets in Cantonese, which rests on ranked, violable constraints, renders such rules ‘unnecessary’. Archangeli (1997: 27) sums up the prevailing view when she says that in OT, ‘there simply is no rule component at all’. Hammond (1995: 6) elaborates on the idea that OT is rule-free: after claiming that in OT ‘rules are virtually gone’ he goes on to say that ‘The change performed by rules is factored out into a single operation [our emphasis] termed GEN’. This clearly suggests that GEN is functionally very different from, and considerably simpler than, rules. Yet there is no explanation or exploration of that assumption, either in Hammond’s paper or elsewhere. In
fact, virtually all research in OT to date has glossed over the inner workings of GEN, upon which OT's claim to rule replacement crucially relies.

Our present purpose is to examine whether ranked constraints have indeed made rules redundant, by considering whether and how rules are circumvented within OT. We maintain that when the generation, as well as the evaluation, of candidates is considered, GEN might well be regarded as the new incarnation of rules, and although processes are not overtly discussed and are no longer central to phonological theory, they are not redundant in OT. The claim that OT has done away with rules obscures what is realmente a revised view of phonological rules, particularly with respect to their relationship to constraints in phonological theory, since in OT rules do not apply in response to constraint violations—i.e. they are not repair strategies. Nonetheless, this revision of the role of rules in OT, through GEN, represents not a fundamental change in our view of what a rule accomplishes—which is to relate one level of representation to another—but a continuation of the steady decline in the status and role of arbitrary rules that has been taking place for over two decades.

We will show explicitly, using mainly examples from the adaptation of borrowings, that to generate an optimal candidate, GEN, sometimes with the help of the post-phonological interpretive component, effects the operations to which rules had been reduced prior to the advent of OT. Even under the most restrictive view of GEN, it must be relied upon to insert content, as well as structure, and to delete at least structure. However, we are not suggesting that proponents of OT, most particularly McCarthy and Prince, are unaware of the extent of GEN's powers to alter the input. For although GEN's capacities and limitations are generally implied or, at most, brief stated, rather than being argued for or even demonstrated, they are nonetheless intended, intrinsic aspects of OT.1 The idea behind this lack of argumentation or demonstration is OT's belief that what GEN does (i.e. insert content and structure and delete at least structure) is relatively unimportant, or beside the point, because OT places such a heavy burden of phonological explanation on the constraint system.

While we agree with OT's general view of the balance of power between rules and constraints, the fact remains that as long as OT retains GEN in its current conception, it is a procedural theory. To wit, OT is based on the assumption of a representation, which comprises the input to GEN, which can be modified by the addition or deletion of any amount of content or structure, to create a set of potential output forms, which OT calls candidates. The actual surface form is simply the particular output of GEN which best satisfies the constraint hierarchy.

As just hinted, a notable feature of OT, that distinguishes it from either Harmonic Phonology or TCRS is that GEN effects the changes to the input prior to encountering the constraints, rather than as a response to the constraints.

---

1 There have been proposals (e.g. Golston 1997) which would reduce or even gut GEN, but such views have not yet been incorporated into what Golston refers to as 'Standard OT'.

However, this does not strike us as grounds for discounting the changes that occur. The point, it seems to us, is that they occur at all. Indeed, we see this as a fundamental paradox in OT: it repudiates the notions of being procedural or derivational, but it nonetheless relies on the operations that defined rules to transform an input form into an output surface form.

The remainder of the chapter is organized in the following way: Section 6.2 provides a brief overview of the form and function of SPE-type rules. We discuss how and why the form of rules changed during the course of the 1980s, as well as how and why defined limits were set on their power. We also consider the emergence of constraints and how they came to occupy a pre-eminent, if generally non-formal, place in generative phonological theory. A summary of these discussions is provided in 6.2.3.

In Section 6.3 we turn to OT and its handling of the traditional rule functions of insertion and deletion of content and structure, applying these notions to the adaptation of French loanwords in Fula. Loanword adaptation is then used to demonstrate the insufficiency of parsing and non-parsing, the mechanisms first employed by OT in lieu of insertion and deletion of content and structure. Section 6.4 takes up the issue of recent innovations to GEN under Correspondence Theory. The conclusion is presented in Section 6.5.

### 6.2 The Decline of SPE-Type Rules

Claiming victory over SPE-type context rules, in the 1990s, might be like claiming victory over a retired, if not quite dead, horse, because generative phonologists had been doing without SPE-type rules for some time before constraint-based theories, and more recently OT, made their debut. In this section we will briefly sketch the decline of SPE phonological rules and their replacement by rules which, although they still effect the input-change-output cycle inherent in relating one level of representation to another, are less powerful, though more principled, more restricted, and considerably more general in nature than SPE rules. In addition, these more principled rules are, in some frameworks at least, functionally motivated rather than being arbitrary, as in SPE.

The core purpose of a phonological rule has always been to relate one level of representation to another. They have been used to model changes in, and alternations among, phonological representations. That view of their task has often been made explicit, as these representative statements indicate:

The phonological system of a language is comprised of two parts, representations and rules, and the relations between representations (i.e. rules) are described by [Archange and Pulleyblank 1994: 283].

Formal rules characterize how underlying and surface representations are related. (Archange and Pulleyblank 1994: 285)

The phonological rules, then, make up the phonological component of the grammar, and
their function is to convert the UR [underlying representation] of any utterance into its corresponding FR [phonetic representation]. . . (Kenstowicz and Kisseberth 1979: 32)

The two representations [underlying and surface] are systematically related by phonological rules that delete, insert, or change sounds in precise contexts. (Kenstowicz 1994: 7)

. . . phonologists have couched alternations in terms of a process applying to one type of representation to derive another representation. (Spencer 1996: 45)

As some of these statements might suggest, the debates spawned by research into phonological rules came to focus primarily on identifying the levels that were to be related, and on what a rule could and could not do to relate one level of phonological representation to another.

In early generative phonology (following SPE), the specific job of rules was to assign a phonetic representation to the syntactic surface structure. More important than this view of their specific task was the assumption that the phonological component was synonymous with the system of rules. For some time after their introduction within generative phonology, rules constituted the only available explanation for phonological changes and alternations: a given alternation occurred because the language had a rule to that effect. For example, the fact that, in English, the vowel /i/ is realized as [i] in a word like companion was attributed to the rule in (1) (SPE: 87):

(1)  i → j / [dental] C + V

The basic assumption of the early generative view was that, in relating one level of representation to another, rules were capturing the phonological generalizations of the language. The rules were seen not only to express the 'how' of phonological phenomena in a language, but also to stand intrinsically for the 'why' of those phenomena.

6.2.1 Generalizing and Limiting SPE-Type Rules

As research into the identification and elaboration of phonological rules proceeded, phonologists became increasingly aware of the need to shift their focus away from the rules to the phonological representations. That is to say, the nature of the input and the output, not just the processes to which they submitted, came to assume greater importance. Very soon theories of representation not only augmented rule-based models but took priority, as indicated by the familiar slogan 'if the representations are right, then the rules will follow' (McCarthy 1988: 84). In a broader perspective, the shift in focus from rules to representations led to a revised view of where the real generalizations in phonology lay. Autosegmental phonology and feature geometry, in particular, were motivated by the view that representations rather than rules were the bread and butter of phonology. During the 1980s this conviction launched a research program within those frameworks that led to a radical revision of rule form and a drastic curtailment of what was widely acknowledged to be the excessive power of SPE-type rules. More specifically, as a result of work within autosegmental phonology and feature geometry, phonological rules were reduced to basic operations of insertion or deletion, mostly of association lines (Archerelli and Pulleyblank 1986; see also McCarthy 1988 and Goldsmith 1995 for an overview). Nasal place assimilation provides a good illustration of the changes in rules that occurred. Under SPE's approach, nasal place assimilation is expressed by the following rule (McCarthy 1988: 86) which basically states that a nasal takes on whatever values for the features [coronal], [anterior] and [back] are possessed by the following consonant:

(2)  Place assimilation in SPE

[+nasal] → [α coronal] / [β anterior] / [γ back]  →  [α coronal] / [β anterior] / [γ back]  where α, β, γ are variables over + and –

Among other deficiencies of this approach, the formalism offers no insights into why the language, indeed many languages, have this process, as opposed to other processes that are formally conceivable, though unusual or even unheard of.

In contrast, feature geometry reduces the rule of nasal place assimilation to the insertion of an association line, as shown in (3):

(3)  Place assimilation in feature geometry (McCarthy 1988: 87)

Nasal | Consonant |
| o Place |
| . . . |

In addition to streamlining the rule of nasal assimilation, feature geometry was able to offer a deeper level of explanation for the process than SPE. Recall that in SPE the rule, in and of itself, is the inherent reason for change, and there is no formal explanation for why a language has a particular rule. Within the framework of feature geometry, a deeper level of explanation is integrated into the expression of the rule. To return to the example of nasal place assimilation, it was hypothesized to occur because dental/alveolar nasals lack place features of their own. The (widespread, if not totally universal) lack of place features for anterior coronals inclines them to assimilate in place to a following segment specified for place (cf. Paradis and Prunet 1991 for an overview of the special properties of coronals and their relationship to coronal underspecification).

As just mentioned, by focusing on representations and pursuing theories of autosegmental phonology and feature geometry, phonologists addressed a serious
familiar concrete example. As shown in (4), the initial obstruent of a word is voiced when the word is the second member of a compound (Kenstowicz 1994: 162):

(4)  

iro ‘color’ + kami ‘paper’ → origami ‘colored paper’

However, as seen in (5), the initial obstruent is not voiced if there is already a voiced obstruent, here 2, within the word.

(5)  

kami ‘divine’ + 2azi ‘wind’ → kannigaki ‘divine wind’ (*kanigaki)

This was explained as avoiding a violation of the Obligatory Contour Principle (OCP), an independently attested constraint prohibiting adjacent identical featural specifications (in this case, +voice).

(6)  

| kazi  

| gazi  

| [ +voice ]  

| [ +voice ]  

| [ +voice ]

Yip (1988) provided an in-depth analysis of both the triggering and blocking effects of the OCP and showed that having rules that thus serve constraints allowed a large class of rules to ‘be stated as context-free insertion and deletion rules’ and rendered alpha notation unnecessary (Yip 1988: 65). In this way, focusing on constraints led to a further erosion of traditional phonological rules.

At least two constraint-based theories that predate OT—i.e., TCRS and Harmonic Phonology—crucially incorporated context-free rules that were, in essence, responses to constraints and/or well-formedness conditions. In TCRS, context-free rules are called Repair Strategies. Their expression function is to transform input representations that are deficient with respect to the requirements of particular constraints into output representations that meet the well-formedness criteria established by those constraints. They do so by inserting or deleting content or structure. In other words, the ‘rules’ of TCRS are processes which are a) operationally restricted (according to the limits imposed on rules by autosegmental phonology and feature geometry) and b) context-free, since they are motivated entirely by constraints which include both principles (universal constraints) and parameter settings (language-specific answers to universal options). Indeed, repair strategies have very little to do with the languagespecific contextual rules of SPE.

6.2.3 Summary

To sum up, the traditional view that the phonological generalizations of a language lay in language-specific context rules and rule systems was superseded by the view that the real generalizations lay in the output or, more precisely, in the restrictions and preferences that outputs and output alternations revealed. To the extent that rules functioned to bring about a particular condition in their out-
puts, they came to be regarded as mechanisms for enforcing conformity to constraints, rather than as a driving force in and of themselves.

The focus on phonological representations led to much stricter limits on the power and form of phonological rules, while the focus on constraints led to a revised view of the *raison d'être* for rules, and to the removal of their statements of context. Even before constraint-based frameworks received formal articulation, three fundamental and interrelated changes took place in generative phonology: (1) our view of the phonological component implicitly broadened to include constraints, as well as rules; (2) the burden of explanation for phonological change steadily shifted from rules to representations to constraints; and (3) the importance of rules was successively downgraded.

By the end of the 1980s, the glory days of phonological rules had definitely passed, and it was rare to find SPE-type rules used anywhere outside introductory courses in generative phonology. Yet throughout these changes, which saw fundamental alterations in their form, a severe curtailment of their power, and a precipitous decline in their status, the basic function of rules—noncontextual as well as contextual—went essentially unchanged and unchallenged: they remained a mechanism for relating one level of phonological representation to another.

### 6.3 OT and the Insertion and Deletion of Content and Structure

The general aim of this section is to show how *GEN* is used to accomplish what was previously handled by rules—i.e. relate one level of representation to another through the insertion and deletion of content and structure. We will see how, under the most restrictive view of *GEN* that has been taken in standard OT, *GEN* achieves the effects of insertion (6.3.1) and deletion (6.3.2). In 6.3.3 we will show how these mechanisms apply to loanword adaptation where the focus of insertion and deletion is generally at the featural level rather than the segmental as in the cases used to introduce and refine OT. In 6.3.4 it will be shown that *GEN* must perform the traditional rule functions of insertion of content as well as structure, and deletion of structure; it cannot achieve the desired results simply through parsing and unparsing.

#### 6.3.1 Insertion of Structure and Content

In the earliest days of OT, *GEN*'s function was to create candidates primarily via the insertion of structure. The effect of inserting content was achieved through the collaboration of a post-phonological phonetic component. The implication was that *GEN* inserted structure, while the post-phonological phonetic compo-

---

1. McCarthy and Prince (1993) assume the widely accepted view of reduplication, that a reduplicative morpheme is an element of the input, designated *Ren*, whose precise phonological form is underarticulated and thus dependent on that of the element to which it is affixed. Reduplication has provided fertile soil for OT; many crucial aspects—such as whether reduplication is total or partial, suffixed or prefixed, whether elements are inserted or deleted—are attributed to the ranking of particular constraints.
ing from V+V or C+C sequences are avoided in the output forms, as shown in (5).

(5) Input  Optimal candidate  Phonetic output
       a. ir-saik-i  i(r)saiki  isaiki  'will sit'
       b. no-ana-ni  n(o)anani  nanani  'my black dye'

In (5), the candidates that satisfy the constraints against V+V or C+C sequences leave r and o, respectively, unparsed. It is by virtue of leaving these segments unparsed that these candidates best satisfy the constraint hierarchy and are submitted for phonetic implementation. Because r and o are unlinked to higher prosodic structure, they are phonetically deleted. OT maintained that input material was not actually deleted, but if it is left unlinked, it cannot be realized phonetically.

6.3.3 Applying Parsing and Non-Parsing to Loanword Adaptation

McCarthy and Prince (1993) and Prince and Smolensky (1993) focus on the parsing and non-parsing of whole segments and segmental positions, but loanword adaptation can be used to illustrate these operations at the level of distinctive features. If we take the position that loanword adaptation is phonological (Hyman 1970; Prunet 1990; etc.), which has most recently been extensively argued by Paradis and LaCharité (1996, 1997), an unadapted loanword (the word as it occurs in its language of origin) and an adapted loanword (the word as it occurs in the borrowing language) would constitute in standard OT, respectively, input to, and output of GEN that are directly observable.

Loanwords are particularly interesting for two reasons. First, the nature of loanword input does not have to be hypothesized to the extent that native word input often does. Second, an unadapted loanword often contains what are, from the point of view of the borrowing language, felicitous feature combinations and structures. That is to say, in their original state, loanwords are often an affront to the structural and segmental constraints of the borrowing language. It is for this reason that loanwords are, whenever necessary, adapted through the apparent addition or subtraction of features and/or structure. Of particular interest about loanwords is that, irrespective of the particular choice of features assumed for segments, and of one's position on various issues, such as whether input representations in OT are underspecified or not, whether features are binary or monovalent, etc., the fact remains that the optimal candidate for a loanword is often quite unfaithful to the input.4

Put another way, an adapted loanword must incur violations of faithfulness constraints in order to satisfy what are obviously more highly-ranked segmental and structural constraints of the borrowing language. Informally characterized, faithfulness constraints require that output form, in this case the adapted loanword, contain all and only the material of the input form, in this case the unadapted loanword.6

Let us now consider in more detail how the adaptation of loanwords would have been handled in early versions of OT, which considered GEN capable only of inserting structure and leaving material unparsed. Recall that OT considered that the 'input to the grammar is a string of root nodes each with a set of (unassociated) features. The output is an optimal parse in which these features are associated to root nodes (with the root nodes associated to syllable position nodes, and so on up the prosodic hierarchy)' (Prince and Smolensky 1993: 180). This meant that GEN assembled the featural input into segments and the constraints of the borrowing language were responsible for selecting the candidate that represented the best assemblage. Consider how this works in the case of the French word marchandise [marʃɑ̃dʁɛs] 'merchandise', which is borrowed into Fula as [məɾsəndis]. Fula is a West African language that has neither the alveopalatal fricatives nor the nasal vowels that occur in French. Of particular focus here are the adaptations of the alveopalatal fricative [ʃ] and the following nasal vowel [ə], neither of which are licit segments in Fula. Let us assume, for the sake of illustration, that the constraints barring these segments are those in (6a) and (6b):

(6) a. *[ʃ]: *[+continuant] [−anterior]
b. *[ə]: *[+consonantal] [+nasal]

The constraint in (6a) is intended to indicate that the co-occurrence of the features [+continuant] and [−anterior] is prohibited in Fula. The constraint in (6b) indicates that nasal cannot occur on vowels. By virtue of containing the alveopalatal strident [ʃ] and the nasal vowel [ə], the unadapted loan violates these Fula constraints. The fact that the word is adapted indicates that these structural constraints are higher-ranked than the faithfulness constraints which would have the segments remain unchanged.

From an OT perspective, GEN creates from the input a set of candidates. The ranked constraints of Fula will select that candidate which it deems best. Let us consider what it means to come up with the optimal candidate for a word which, in its unadapted state, is an unacceptable surface form in Fula. In this case, the optimal candidate, [məɾsəndis] is created by, among other things, parsing the feature [+continuant] and leaving the feature [−anterior], which co-occurs with [+continuant] in the input, unparsed, as shown in (7).

---

4 These authors show that the borrowers and adapters are bilinguals—the term being taken in its broad sense—and provide arguments against the phonetic stance supported by Silverman (1992), for instance.

5 Prince and Smolensky (1993) propose that learners infer input forms that, by-and-large, reflect surface constraint preferences. Violations of faithfulness in underlying native forms are therefore expected to be much less usual than in borrowed words.

6 Under Correspondence Theory, the material is further required to serve the same prosodic functions in input and output.
Because [-anterior] is not parsed (i.e., linked to a root node and higher prosodic structure), it is stray-erased at the phonetic level. In this way the deletion of features, as well as of segments and segmental positions, is effected through non-parsing by GEN and subsequent stray erasure at the level of phonetic implementation.

In the case of the nasal vowel, the feature [+nasal] is parsed in the output, but instead of being linked to the vowel, as it is in French, it is linked to both an epenthized coda consonant and to the following onset consonant, to produce the optimal output form [marsan*dis]. This is illustrated in (8) (for clarity, articulated syllable structure is used and the epenthized coda is underlined):

\[
\begin{align*}
\sigma & \quad \sigma & \quad \sigma \\
O & \quad R & \quad O & \quad R & \quad O & \quad R \\
N & \quad C & \quad N & \quad C & \quad N & \quad C \\
marsa & \quad dis \\
\[\text{[-cons]} & \quad [+\text{nas}] \\
\end{align*}
\]

By linking [+nasal] to an epenthized coda consonant, GEN produces an adaptation that meets the requirements of Fula's constraint hierarchy. Those features not present in the input, and which remain underspecified in the output of GEN—for instance the coronal place of articulation for the epenthized coda nasal—are filled in phonetically.

6.3.4 The Insufficiency of Parsing and Non-Parsing

We have just seen that GEN may leave individual features unparsed for post-phonological stray erasure in the same way that it leaves whole segments unparsed, to be stray-erased by the phonetic implementation component. We have also seen that just as GEN inserts consonant and vowel slots to be filled in by default features phonetically, so it may insert consonant or vowel slots to host features that are differently distributed in the unadapted loanword. In short, to the limited extent that we have pursued the topic, loanword adaptation does not seem to tax OT, or to require any additional assumptions. Even if GEN is restricted to parsing and non-parsing, it appears equal to the task of loanword adaptation.

However, a deeper look at the adaptation of loanwords reveals that such a restricted view of GEN is untenable. In addition to leaving features and segments unparsed, or linking them in different ways than occurs in the unadapted loanword surface form, GEN must also be able to insert content and to delete structure. We consider, first, the insertion of content by GEN. The possibility that GEN might be required to insert content, as well as structure, was clearly entertained by Prince and Smolensky (1993: 103) who acknowledge that "perhaps, for example, the set of candidates issued by GEN should include actual featural and segmental insertions, as well as new association lines." Loanword adaptation shows this to be the case. Consider the adaptation of *v that occurs in French loanwords borrowed into Fula, which lacks voiced fricatives, and which adapts those that occur in French borrowings. Examples are given in (9a), along with examples of *z and *j adaptation in (9b).

(9) a. The adaptation of *v in Fula loanwords

<table>
<thead>
<tr>
<th>Adaptation</th>
<th>French</th>
<th>Fula</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. v → w</td>
<td>avocat</td>
<td>[ayɔk] → [ayɔk]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cível</td>
</tr>
<tr>
<td>ii. v → b</td>
<td>avion</td>
<td>[ayjɔ] → [abijɔn]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>livre</td>
</tr>
<tr>
<td>iii. v → f</td>
<td>élève</td>
<td>[ɛlɛv] → [elev]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mouvement</td>
</tr>
</tbody>
</table>

b. The adaptation of *z and *j in Fula

<table>
<thead>
<tr>
<th>Adaptation</th>
<th>French</th>
<th>Fula</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. z → s</td>
<td>kerosene</td>
<td>[kerozen] → [kerozen]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>télévision</td>
</tr>
<tr>
<td>ii. j → s</td>
<td>barrage</td>
<td>[baraj] → [baraj]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>chambre</td>
</tr>
</tbody>
</table>

Both the absence of voiced fricatives in Fula words and their adaptation in loanwords (cf. 9b) suggest an undominated phonological constraint against the co-occurrence of [+cont] and [+voice], the feature combination assumed to define voiced fricatives. A priori, an optimal output of GEN may not contain the infelicitous feature combination or it will not satisfy the constraint. To produce a candidate that satisfies this highly-ranked constraint, GEN might leave either feature, [+continuant] or [+voice], unparsed. Failure to parse a feature ultimately allows the non-phonological interpretive component to supply redundant and default features ( McCarthy and Prince 1993: 24), in some cases replacing the feature value of the input. For instance, in the case of *v, if [+voice] were left

7 The relevant Fula constraint cannot be as specific as *{+strident} [+voice] because it must also account for the adaptation of *v (cf. 9a). It has been shown (Lahiri and Evers 1991; Shaw 1991; LaCharité 1993; Rubach 1994; Steriade 1994) that [strident] is a coronal dependent, and [v] and [f] cannot be considered coronal.
unparsed, the interpretive insertion of [+voice] would yield adaptation to [f]; if [+continuant] were left unparsed, the interpretive insertion of [-continuant] would yield adaptation to [b]. Both of these attested adaptations are schematized in (10):

(10) Input to Gen:

- [+voice] Labial [+cont]
  
  Root node
  
  Unlinked features

Output of Gen:

a. [+voice] unparsed

- [+] Labial [+cont]

  - [+voice] Labial [+cont]

  Phonetic implementation:

a. [+voice] unparsed

- [+] Labial [+cont]

b. [+continuant] unparsed

- [+] Labial [+cont]

  - [+voice] Labial [+cont]

  stay-erased

  ↓

  [f]

However, although both the adaptation of French v to b and f in Fula are well attested, neither is Fula's preferred adaptation. In our data, [v] is adapted to [b] in only 14 cases out of 8i, i.e. in 17.3 per cent of the cases; [v] is adapted to [f] in only 5 cases out of 8i, i.e. in 6.2 per cent of the cases. As in many other languages, v in Fula is usually adapted to w, as in the examples in (9a1). In our database, this occurs in 62 cases out of 8i, i.e. 76.5 per cent of the cases. Adaptation of [v] to [w] shows that Gen must insert content; we cannot assume that it inserts structure that is supplied with content only post-phonologically. Specifically, glides are both [+voice] and [+continuant], by virtue of the fact that they are [+sonorant]. This suggests that in the adaptation of *v to w, [+sonorant] is added to [+continuant] and [+voice] to license the otherwise unacceptable combination. However, unlike insertion of the redundant values [+voice] and [+continuant], which produce surface [f] and [b], respectively, [+sonorant]—which can by no means be considered a redundant or default feature in the cases at hand—must obligatorily be inserted by Gen. If Gen did not insert [+sonorant], the candidate could not be selected because it would not satisfy the constraint "[+continuant] [+voice]. In other words, the winning candidate can only be the winning candidate if it satisfies this constraint (among others, of course).

We point out that issues of underspecification are beside the point here. That is to say, the fact that the output of Gen is more than minimally or contrastively specified—since it bears both [+sonorant] and, redundantly, [+voice]—poses no problem for OT, which does not rely on underspecification to anywhere near the extent that previous generative theories have done (cf. Inkelas 1994; Ito et al. 1995). Although the preceding discussion of ependent in Axinica Campa indicated that the interpretive component adds redundant and default feature values, OT does not necessarily bar such features and feature combinations from occurring earlier. Thus, the fact that Gen inserts [+sonorant], which is in this instance redundant, to the feature combination [+continuant] [+voice] to create an optimal candidate from French loanword input containing v is not an issue from the point of view of underspecification. It does show, though, that Gen inserts not only structure but content as well.

One might be committed to maintaining a non-procedural view of Gen in which Gen only parses or leaves unparsed, and is unable to insert content not present in the input. From this perspective, an alternative might be considered which relies on candidate outputs where either [+voice] or [+continuant], or both, are simply left unparsed. Then, [+sonorant] insertion (like [+voice] or [+continuant] insertion that produce adaptation to f or b, respectively) is left to the phonetic interpretive component. As with the f and b cases illustrated in (10), such outputs of Gen could then be selected as optimal because they would not really instantiate the offending feature combination. However, any such analysis faces a significant problem: there is no principled means for ensuring the phonetic insertion of [+sonorant], as opposed to some other feature. For instance, why would a labial, in a consonant position, which emerged from Gen without continuant or voicing specifications, be interpreted as a glide? In other words, if the optimal candidate emerged from Gen with neither [+voice] nor [+continuant] parsed, why would the segment not be interpreted as a more prototypical consonant, such as a voiceless labial stop? Even if the labial emerged with [+voice] parsed, why would the segment not be interpreted as a voiced labial stop, rather than as a glide? If the emergent candidate contained the features Labial and [+continuant], why would not be interpreted as a voiceless labial fricative, rather than as a sonorant? In short, no matter which feature we choose to leave unlinked in order to avoid a fatal constraint violation, [+sonorant] is a very unlikely feature to be inserted as a default or redundant value for a consonant. We conclude that [+sonorant] must be inserted by Gen.

* The remainder of the cases involve metathesis (9 cases, i.e. 9.1 per cent) and deletions (9 cases, i.e. 9.1 per cent), which occur under predictable phonological conditions (Paradis and LaCharité 1997).
Now let us reconsider deletion by GEN. By handling deletion via the mechanism of unparsing, OT implied that GEN did not delete association lines. However, the adaptation of loanwords again shows that this restrictive view of GEN does not work. GEN must be able to delink—that is, to delete structure, as well as insert it. Although segments in inputs may be ‘unassembled’, McCarthy and Prince (1993: 21) contend that ‘Vowels, long or short, come with moraic structure attached in the lexicon’. Therefore, if vowels in loanwords are shortened or consonantalized, to produce an optimal candidate, then GEN must have delinked the vowel from its mora. In fact, McCarthy and Prince (ibid.) allude to this possibility when they say that ‘underlying vowel length distinctions are represented by lexical mora specifications, so they must be present (though not necessarily realized) in all candidate forms’. Examples of delinking of vowels from their moras are provided by French loanwords borrowed into Kinyarwanda.

In each of these examples, part of a diphthong becomes part of an onset (cf. e.g. Kaye and Lowenstamm 1984 for a discussion of diphthongs in French and arguments). That is to say, in adaptation, a moraic glide becomes a non-moraic on-glide. Let us consider French [fraswa] ‘Francois’. Because [wa] is a diphthong in French, [w] must be linked to a mora. However, both diphthongs and nuclear segment sequences are unacceptable in Kinyarwanda, and must be adapted. In the examples in (11), we see that, in the Kinyarwanda adaptations of these French loanwords, [w] labializes the preceding consonant, indicating that it is linked to the onset, clearly a non-moraic consonant position. This means that [w] must delink from its mora. In cases where the moraic glide cannot become an on-glide, as with onset fricatives, which cannot be labialized, not only is the glide delinked from its mora but it is totally lost, as shown by the examples in (12).

The changes exemplified in (13) are particularly reminiscent of the delinking and spreading operations that were so characteristic of autosegmental phonology and feature geometry. We conclude that such loanword adaptations show that GEN can delink and spreading features, or more to the present point, that GEN can delink features as well as simply leave them unparsed.

To summarize, OT as it was first conceived by McCarthy and Prince (1993) and Prince and Smolensky (1993) allowed for the insertion of content and structure and the deletion of content and structure, though not in the conventional manner of rules. In OT, primary responsibility for inserting structure rests with GEN, which creates a candidate set for assessment by ranked constraints. The primary means of effecting deletion was through failure to parse particular material, leading to its phonetic deletion. However, applying OT to loanword adaptation shows that the desired effects cannot be achieved simply through parsing or failing to parse. GEN’s operations are shown to include at least those in (14):

Both the insertion of content (cf. the discussion of adaptation of French [v] to [w] in Fula, via the insertion of [+sonorant] in 6.3.4) and the deletion of structure (as just shown by the cases where French diphthongs are delinked from one of their moras in Kinyarwanda adaptations) cast doubt on the claim that OT is non-procedural, and that the relation between input and output is

---

* Although complex consonants—like labialized and palatalized consonants—are permitted in Kinyarwanda, complex (i.e. branching) onsets, codas or nuclei (except for long vowels) are not.

---

"That material is unparsed, as opposed to being phonologically deleted, is simply a claim made by OT. It has not, as far as we are aware, been argued for anywhere in the OT literature."
'straightforwardly monotonic' (Prince and Smolensky 1993: 25). Monotonicity depends crucially on the idea that GEN in effect just provides different possible ways of looking at the input—that it simply interprets the input, so to speak. If GEN actively creates output candidates by inserting content, and more especially, by effectively deleting content through the erasure of structure, then the relation between input and output cannot be considered either monotonic or non-procedural.

6.4 Lifting the Restrictions on GEN

OT has undergone a recent important refinement, referred to as Correspondence Theory, which has had an impact on the characterization of GEN (cf. McCarthy 1995). The purpose of this section is to provide a brief look at the main changes to OT under Correspondence Theory in order to show that they do not obviate the point that we raise here, which is that GEN is a procedural component. Indeed, under Correspondence Theory, GEN is more procedural, not less so.

The crux of the change to OT under Correspondence Theory is that faithfulness is achieved somewhat differently. The faithfulness constraints Parse and Fill have given way to a more articulated, more powerful and more general class of what are now called correspondence constraints. One difference between the implementation of faithfulness using Parse and Fill and the implementation of faithfulness in Correspondence Theory is that in the latter, constraints press not only for substantive identity between input and output, as the faithfulness constraints Parse and Fill do, but they further require that segmental content in each member of the pair serve the same prosodic function. For example, a segment that functions, in the input, as the nucleus of a syllable is required by the (relevant) correspondence constraint to function as the nucleus of a syllable in the output. Thus, delinking a segment from the nucleus and relinking it to the onset, as occurs in the Kinyarwanda adaptations shown in (11), is a violation of the correspondence relation, even though the segmental material is otherwise unaltered.

Another difference between faithfulness and correspondence is that the correspondence relation governs not only the underlying and surface pair, as the faithfulness constraints were conceived to do, but surface alternations as well. For instance, McCarthy (1995) shows how correspondence constraints govern the incomplete and complete phase forms in Rotuman, both of which are surface forms. However, this proposal does not have a bearing on the point being made here, mainly because no loanword that undergoes adaptation could be related to its unadapted counterpart in this way. In other words, an unadapted loan containing illicit segments or structures is an unacceptable surface form in the borrowing language. To be a surface form, it must be adapted, which means that it must pass through GEN. The pertinent question remains, then, whether GEN is any less procedural under Correspondence Theory. We suggest that it is not.

In fact, under Correspondence Theory, GEN is much less restricted in its operation than it was initially conceived to be, and OT no longer maintains that GEN is restricted to effecting changes between input and output solely through parsing and non-parsing. The current position is that 'GEN is quite creative, being able to add, delete, or rearrange things without restriction' (Archangeli 1997: 14). Lifting the restrictions on GEN is partly attributable to the fact that as research into the identification and articulation of constraints has continued in OT, the need to restrict GEN's operations to parsing and non-parsing has been rendered unnecessary because the constraint system has emerged strong enough to deal with whatever GEN might come up with. However, the restrictions on GEN have not been relaxed solely because they are unnecessary; they have necessarily been lifted. In early OT, GEN was governed by three principles, one of which was Containment. Containment required the input to be contained in every output, something that is at odds with the notion of deletion (cf. McCarthy and Prince 1993: 20). However, Containment is not only unnecessary under assumptions of Correspondence Theory, it restricts constraint ranking in ways that contradict actual outcomes (see, for example, McCarthy 1995: 13) and is thus antithetical to Correspondence Theory. So Containment has been abandoned, and along with it, the restriction of GEN's operations to parsing and non-parsing. The net result is that while OT would formerly have resisted the idea that GEN inserts and deletes content and structure, OT under Correspondence Theory openly accepts that it does. Therefore, if GEN was already procedural in early versions of OT, there can be no doubt that it is equally if not more procedural now.

6.5 Conclusion

Prior to the emergence of formal constraint-based theories, the power and role of phonological rules had already been seriously eroded. In a de facto sense, if not always formally, the power of rules has been limited to inserting or deleting content or structure (see Section 6.2.1). Rules were increasingly treated not as an explanation for phonological changes and alternations but as the servants of phonological constraints, wherein the real explanations were seen to lie (see Section 6.2.2). When rules ensure conformity to constraints, they are also context-free.

OT is a highly articulated effort to account for phonological regularities in a language in terms of constraints, rather than rules. Indeed, OT is billed as a rule-free alternative to derivational constraint-based theories (Prince and Smolensky 1993: 5) such as TCRS and Harmonic Phonology, which indeed resort to 'rules'. The rules they resort to, though, are universal and rely on a violated (universal or universally-based) constraint to provide context, and their power is limited to insertion/deletion of content/structure. TCRS terms such operations 'repairs', but OT considers them rules, which is often misleading. This view is
obvious in Roca (1997: 14), who, noting the seminal difference between OT and TCRS, points to the latter's use of rules. The reliance versus non-reliance on repair is thus considered to be a defining point of difference between OT and TCRS. OT's basic claim might well be characterized as 'take care of the constraints and the rules become unnecessary'. However, given the revolutionary nature of OT's position, it is crucial to probe the mechanisms by which disposition of phonological rules is to be accomplished. It is insufficient to simply claim that 'GEN . . . generates for any given input a large space of candidate analyses by freely exercising the basic structural resources of the representational theory. The idea is that the desired output lies somewhere in this space . . .' (Prince and Smolensky 1993: 5).

As noted in the introduction, OT researchers have thus far been preoccupied with the selection of one candidate from a field of potential output representations, not with the operations behind the provision of candidates. Because of that, OT has been able to maintain—sparsely, we think—that it has done away with rules. However, if in order to generate that field of candidates, or even one optimal candidate, GEN, even in its most restricted conception, must have the power to insert content as well as structure, and to delete structure, then we must conclude that, contrary to its claims, OT does indeed rely on rules, as well as constraints. GEN might then be seen to be a cover-term for the most recent incarnation of the phonology's rule component.

Archangeli (1997: x) likens OT's task to that of a fisherman who turns his attention away from designing the ideal net to designing the ideal separator, in order to weed out whatever undesirables his imperfect net admits. While the analogy is enlightening, there is a crucial difference that we feel cannot be overlooked: the fisherman's net gathers what is already in the environment, whereas GEN actually creates the undesirables which the separator must subsequently deselect. Surely the idea that GEN is able to insert or delete any amount of content or structure casts doubt on the ideas that GEN constitutes a single operation (cf. Hammond 1995: 6), that 'There are no rules or repair strategies, and no serial derivation', (cf. Ito, Mester and Padgett 1995: 578), that an OT account renders rules 'unnecessary' (Yip 1993: 2), or that GEN really dispenses with the structural change of rules (Prince and Smolensky 1993: 5). OT cannot disavow the use of repair while depending on the operations that define repair (it would be trivial, and dishonest, to reject the mechanism but not the operations that define it). If our interpretations of loanword adaptations within OT are correct, or even if they are to be meaningfully challenged, we must seriously question the idea that GEN functions in a radically different way from the way that rules do.

Regardless of the fact that OT has profitably shifted its focus from the procedures expressed by rules to the constraints and, more important, to constraint systems, the fact remains that if GEN is a rule component by another name, then one of OT's basic claims is compromised. Roca (1997: 33) says that OT can claim a formal advantage over derivational constraint-based theories such as TCRS and Harmonic Phonology, insofar as OT does not need rules. If, as we suggest here, this is not the case, then this formal advantage is lost and the real difference between OT and other derivational constraint-based theories may be seen to lie in view of the relationship between rules and phonological constraints.

6.6 References


— and D. Pulleyblank (1986). 'The structure and content of phonological representations'. Ms, University of Arizona at Tucson and University of British Columbia, Vancouver.


Inkelas, S. (1994). 'The consequences of optimization for underspecification'. Ms, University of California at Berkeley.


— (1993). 'Phonological constraints, optimality and phonetic realization in Cantonese'. MS, University of California at Los Angeles.