CHAPTER 3:

THE UNDERLYING FORM – SURFACE FORM CORRESPONDENCE

Between underlying and surface representations, some pieces of structure are the same and are arranged in the same way, held constant by the grammar; other pieces fail to be so, being modified by the grammar. That is, there is a map between the underlying structure and surface structure, which we may call the **underlying-surface correspondence**.[1] In this chapter we compare the two formal approaches to this map: in derivational theory, the underlying form is mapped to the surface form by the operations of rules; in optimality theory, the underlying form is mapped to the surface form by the input-output correspondence relation, which may register violations of Faithfulness constraints.

We show that the complexity of the derivational approach is greater, because some steps of a derivation may be obscured in the overall mapping, but it is always possible to construct a derivation whose overall mapping is **representative** of all its steps. We then identify a more restricted natural class of **veritable** mappings, showing how this property is approximated by the **minimality of constraint violation** metric in Optimality Theory (Pulleyblank and Turkel 1997), and by the principle of **economy of derivation** in Minimalist derivational theory (Calabrese 1995). Finally, we use this formal analysis to explain phonologists' long-standing suspicion of the so-called Duke of York gambit, and we clarify the issue with fresh arguments that the Duke of York gambit is both unexplanatory in general and contrary to the empirical evidence.

---

[1] We assume, after Goldsmith (1976), Coleman and Local (1991) and Bird (1995), that the structures concerned are *graphs* with a "multi-linear" arrangement of several linear sequences (*tiers*) of *nodes* specified with features or other phonological units, and ordered *associations* between nodes on different tiers.

3.1 Rule Operations and Faithfulness Constraint Violations

The basis for a systematic comparison lies in the fact that the types of operations used in structural changes in a derivation and the violations of Faithfulness constraints in an evaluation system may be identified with the same basic set of formal mappings. When a Deletion operation occurs in a derivation, there is some piece of structure in one representation (the **focus**) which is not mapped to anything in the subsequent representation. Similarly, a violation of the constraint MAX occurs precisely when an input-output correspondence contains a piece of structure in the input that is not mapped to anything in the output. So an essential analogy exists between Deletion and MAX violation, for in both instances, there is some structural object that is not mapped to any object. This is a matter of mathematical fact, a breach of the relational property of *totality*. Totality holds when *every* object in the domain is mapped to some object in the range. Similar analogies exist between other rule operations and Faithfulness constraints:

(1)

Disparity	Rule Operation	Faithfulness Violation
$x \rightarrow \varnothing$	Deletion	MAX violation
$\varnothing \rightarrow x$	Insertion	DEP violation
$x_1 \rightarrow x_1 x_1$	Fissure	INTEGRITY violation
$x_1 x_2 \rightarrow x_{12}$	Fusion	UNIFORMITY violation
$x \rightarrow y$	Change of Value	IDENT violation
$xy \rightarrow yx$	Change of Order	LINEARITY violation

Furthermore, any disparity, whether it is viewed as the application of some operation, or as the violation of some Faithfulness constraint, essentially boils down to a breach of one of the stock properties used to describe the behaviour of mathematical relations, defined in (2) (Partee, ter Meulen and Wall 1990:27ff):

(2)

Disparity	Property Breached	Property holds when...
$x \to \varnothing$	totality	every object mapped to some object
$\varnothing \to x$	surjectivity	every object mapped from some object
$x_1 \to x_1 x_1$	function	no object mapped to more than one object
$x_1 x_2 \to x_{12}$	injectivity	no object mapped from more than one object
$x \to y$	identity	every object mapped to an identical object
$xy \to yx$	structure preservation	all objects mapped to objects with same order as the original

Thus, the analogy between rule operations and Faithfulness constraint violations is grounded in the theory of mathematical relations.

3.2 Input-Output Correspondence and Derivational History Relations

The next question is how the disparities combine into an overall mapping. In this section, we consider how this is done in each framework, demonstrating that the complexity of the derivational framework is greater.

3.2.1 Input-Output Correspondences

In an evaluation system, we know from chapter two that for a given underlying form I, all the candidate-triples $\langle I,O,C \rangle$ (where O is any structure and C is any correspondence relation between I and O) are in principle subject to evaluation, and that the system delivers one optimal candidate ⟨I,O_{opt},C_{opt}⟩. Thus, in an evaluation system, the underlying-surface correspondence is an input-output correspondence - specifically, the input-output correspondence of the optimal candidate, C_{opt}.

As well as all other structures O_{subopt} being ruled out, all the other possible correspondence relations for O_{opt} are ruled out as the underlying-surface correspondence. That is, a candidate ⟨I,O_{opt},C_{subopt}⟩ which contains the same output form O_{opt} but assumes a different correspondence relation C_{subopt} will be non-optimal and thus C_{subopt} is not the underlying-surface correspondence. So taking the simplest case when the surface representation is the same as the underlying representation, the evaluation selects the identity relation rather than some more complex mapping. This is shown in (3) for the mapping /bi/ → [bi].

(3)

/bi/	MAX	DEP	IDENT	LINEARITY
a. Input: /b_1i_2/ ☞ Output: [b_1i_2]				
b. Input: /b_1i/ Output: [b_1i]	*	*		
c. Input: /bi_2/ Output: [bi_2]	*	*		
d. Input: /b_1i_2/ Output: [b_2i_1]			**	*

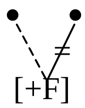
Hence, in the optimality framework, the underlying-surface correspondence is none other than the input-output correspondence of the optimal candidate. It is selected by evaluation against the constraints.

3.2.2 Derivational History Relations

In the derivational framework, an underlying-surface correspondence is constructed from a whole series of structure-to-structure mappings in a derivation. The **serial composition** of these mappings is a construct that matches in form the input-output correspondence of an evaluation.

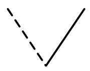
Consider for example a case where a feature spreads (indicated in (4) by the dotted line) leftwards to a new anchor node and delinks (indicated in (4) by double-crossed line) from the original anchor node:

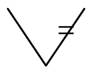
(4)



A concrete instance of this is the Klamath alternation presented in 1.2, where lateral articulation is realised on one segment to the left of that where it is given underlyingly: /n-l̥/ sequences become [lh] and /nlˀ/ sequences become [lʔ]. This is a leftward shift in the association of a [+lateral] feature.

The usual notations for spreading and delinking (association lines dotted or crossed out, respectively) are useful conflations of the relation between successive structures in a derivation. (4) abbreviates a derivation in which there are three stages: between the first and second stages, an operation applies inserting an association between [+F] and the first anchor (•); between the second and third stages, an operation applies deleting an association between [+F] and the second anchor. (5a.) illustrates the insertion (spreading), (5b.) illustrates the deletion (delinking).

(5)(a) insertion		stage 1	stage 2
• •	<=>	• •	• •
		/	\ /
[+F]		[+F]	[+F]

(b) deletion		stage 2	stage 3
• •	<=>	• •	• •
		\ /	\
[+F]		[+F]	[+F]

Call each of the two subsequences from stage 1 to stage 2 and from stage 2 to stage 3 a **step**. Call the mapping between stage 1 and stage 2 "$_1$step$_2$", and call the mapping between stage 2 and stage 3 "$_2$step$_3$". These separate steps somehow combine to give the situation summarised in (4) with two underlying/surface disparities. In fact, this is achieved by serial composition.

(6) **Definition: Serial Composition** (Partee, ter Meulen and Wall 1990:35)

For sets A,B,C, and relations R between A and B, and S between B and C, the serial composition of R and S, a relation between A and C, is given by:

S o R $=_{def}$ $\{(x,z) \mid \exists y$ such that $(x,y)\in R, (y,z)\in S\}$

The serial composition of two mappings relates elements in the first structure to elements in the last structure that are indirectly linked via the second structure.

This we can illustrate in (7): the composite relation of the two mappings ${}_{1}\text{step}_{2}$ and ${}_{2}\text{step}_{3}$ matches items at stages 1 and 3 which are indirectly linked via some intermediate element at stage 2 according to the respective relations ${}_{1}\text{step}_{2}$ and ${}_{2}\text{step}_{3}$.

(7) *first step* *second step*

stage 1 to stage 2 to stage 3

• • • • • •

$/_{i}$ ${}_{j}\backslash\ /_{i}$ ${}_{j}\backslash$

[+F] [+F] [+F]

composite relation:

stage 1 to stage 3

• • • •

$/_{i}$ ${}_{j}\backslash$

[+F] [+F]

The association inserted in the first step (coindexed *j*) is present at the final stage but this lacks a correspondent at the initial stage. The association deleted in the second step (coindexed *i*) is present at the initial stage but this lacks a correspondent at the final stage. In this way the composite relation identifies the two associations as having been inserted or deleted in the course of the derivation.

The application of any rule has a characteristic correspondence relation, exhibiting a breach of some relational property according to what kind of operation is applied (as tabulated in (2)). These characteristic "step"-relations may be composed in series to produce long-range correspondences that relate back structural elements at any stage to their antecedents in the first structure. Such relations express **derivational histories**, a notion already named and discussed in the generative phonology literature (Kenstowicz and Kisseberth 1977). While ${}_1step_2$ links the structure at the second stage back to that of stage 1, ${}_2step_3 \circ {}_1step_2$ now links the structure of the third stage back to stage 1. This leads us to a general definition:

(9) **Definition: Derivational History Relation**

Let the structures at stages 1,2,...,n of a derivation be denoted P1,P2,...,Pn. Let the correspondence mapping characteristic of the rule which applies at each step ⟨Pi,Pi+1⟩ be ${}_istep_{i+1}$. The **derivational history relation** of each structure Pi:

a. ${}_1H_2 =_{def} {}_1step_2$

The derivational history relation between the structure at stage 2 and the initial structure is simply the step relation between stages 1 and 2.

b. for i≥3, ${}_1H_i =_{def} {}_{i-1}step_i \circ {}_1H_{i-1} = \{\langle x,z\rangle \mid \text{for some } y, \langle x,y\rangle \in {}_1H_{i-1} \text{ and } \langle y,z\rangle \in {}_{i-1}step_i\}$

The derivational history relation between the structure at stage i *and the initial structure links elements in the structure that are linked indirectly via the derivational history of the structure at stage* i-1 *and the step from* i-1 *to* i.

Equivalently, the derivational history relation ${}_1H_i$ can be viewed as the serial composition of all the successive step-relations for the steps between stage 1 and stage *i* as given in (10).

(10) i=2,...,n, ${}_1H_i =_{def} ({}_{i-1}step_i \circ ({}_{i-2}step_{i-1} \circ \ldots ({}_2step_3 \circ {}_1step_2)..))$

Since the composition is defined step by step, the earliest derivational steps are given in the most nested position (on the right in (10)) and the later derivational steps in the least nested position.

Having defined the derivational history relation for an entire structure, we can recognise the derivational history of a particular entity within a structure (Kenstowicz and Kisseberth 1977). The derivational history of a particular entity *z* within the structure at stage i of the derivation is any ordered pair $<z_1,z>$within ${}_1H_i$ relating *z* back to some object z_1 in the first structure. But if the object owes its presence to an insertion operation, there will be no such statement and its derivational history is null.

Finally, we observe that in a derivation totalling *n* stages, the derivational history relation ${}_1H_n$ relates the last structure back to the first structure. Hence, the underlying-surface correspondence as given by a derivational grammar is the relation ${}_1H_n$ in the particular derivation concerned.

3.2.3 Identifying The Two

In both frameworks we can relate structural components back to those in the underlying structure, either by input-output correspondences or by serial composition of the mappings at each step of a derivation.

A generation system and an evaluation system describing the same function from a set of underlying forms and a set of surface forms may agree with each other as to the underlying-surface correspondences. They do so if the final derivational history relations and the optimal input-output correspondences are identical: if, in every case, ${}_1H_n = C_{opt}$.

They differ, of course, on how these correspondences are constructed. If there are *n* disparities in an underlying-surface correspondence, they will be contained in the input-output correspondence of the optimal candidate. But a derivational history relation with those *n* disparities is composed serially. Each may be introduced step by step in a series of *n* operations. These *n* steps produce the same final structure whatever order they are placed in, and the number of possible orders is the number of permutations of *n*, given by the formula

(12) $n! = n.(n-1).(n-2)....2.1$ e.g. $4! = 4.3.2.1 = 24$

So there is a natural class of *n*! possible derivations which could be responsible for an underlying-surface correspondence relation containing *n* disparities.

3.2.4 Hidden Operations

In derivations specifically, however, some operations may be obscured by subsequent operations and not be apparent in the final derivational history relation. This possibility adds considerably to the class of *n*! natural derivations that could be responsible for a given underlying-surface correspondence.

We pursue this by means of an example. In many dialects of Italian, there is a metaphony process where mid vowels are replaced by high vowels in the presence of a following high vowel. The vowels that are raised are stressed vowels, in the first syllable of the last foot in a word (Calabrese 1995:450). The examples in (13) are from the Veneto dialect. The metaphony process can be detected in the examples (13a), where it affects "tense", or ATR, vowels. However, metaphony is not apparent in the examples in (13b) where the vowel is "lax"/nonATR:

(13) Veneto Italian (Calabrese 1995:446)

a.	védo	te vídi	I see / you see
	córo	te cúri	I run / you run
	tóso	túsi	boy / boys
b.	prɛ́tɛ	prɛ́ti	priest / priests
	mɔ́dɔ	mɔ́di	way / ways

If the vowels in (13b) were to undergo metaphony, the high nonATR vowels ([ɪ,ʊ]) would be produced, universally marked vowels which in many languages would not be realised on the surface, but substituted for by other, unmarked vowels. Thus, in some other Italian dialects the vowels [i,u] are found in this context (neutralising with ATR vowels); one further dialect has [e,o]. In the Veneto Italian dialect, then, it appear that the repair *lowers* the vowels to [ɛ,ɔ]. However, we then end up with an analysis in which the [+high] feature of the following vowel is spread to the head vowel only to be delinked again:

```
(14)   prɛti      →        prɪti      →        prɛti
             V      V            V      V
              \     |             \=    |
             [+high]             [+high]
```

This is an instance of the so-called Duke of York gambit, where one structural change is reversed by the opposite change (Pullum 1976). In fact, in the overall derivational history relation we simply have no trace of these steps. If there were an association present in the surface form but not in the underlying form, this would evidence the insertion of the association since we would have a piece of structure at the last stage lacking a historical antecedent at the first stage, but here no association line is present either in the surface form or the underlying form, so that no operations are implicated at all. The spreading and delinking operations are not apparent in the overall derivational history relation - *it is just as if they had not applied*.

The alternative, as eventually chosen by Calabrese (1995), is to adopt an analysis of this dialect where the steps in (14) are not used, where the metaphony rule is more strongly restricted so as to spread the [+high] feature only to ATR target vowels, and not to nonATR target vowels.

Then the derivational history relation is a true account of the derivation, there being no "hidden" spreading and delinking.

In fact, as we now demonstrate, whenever the entity produced by one operation is itself subjected to another operation, the form of these two operations becomes obscured in the derivational history relation. This is what happens with several kinds of Duke of York gambit, as in (15). We shall examine some other Duke of York gambits, that have a different effect, in 3.3.3.

(15) Obscured Duke of York gambits

- Inserted elements may subsequently be deleted again.

stepwise	∅ →	a →	∅
overall	∅	→	∅

- Fissured elements may subsequently be re-fused.

stepwise	a →	aa →	a
overall	a	→	a

- Value changes may subsequently be subject to a reverse value change.

stepwise	-F →	+F →	-F
overall	-F	→	-F

- Order change may be subject to a reverse order change.

stepwise	ab →	ba →	ab
overall	ab	→	ab

There are other combinations that serve to obscure operations in a derivation. All deletion of objects previously changed by other operations, and all changes to inserted elements have this effect:

(16) a. Value change and Deletion:

stage	stepwise		overall	
1	-F		-F	
2	+F	Value Change		
3	∅	Deletion	∅	Deletion (Value Change obscured)

b. Insertion and Value change:

stage	stepwise		overall	
1	∅		∅	
2	-F	Insertion		
3	+F	Value Change	+F	Insertion (Value Change obscured)

c. Others

	Order change and Deletion:		Insertion and Order change:		
stage	stepwise	overall	stepwise	overall	
1	a b	a b	a	a	
	X	|	|	|	
2	b a	|	a b	|	
	|	|	X	|	
3	a	a	b a	b a	(order changes obscured)

	Fusion and Deletion:		Insertion and Fusion:		
stage	stepwise	overall	stepwise	overall	
1	a b	a b	a	a	
	∨		│	│	
2	c		a b	│	
			∨	│	
3	∅	∅	c	c	(fusions obscured)

	Fissure and Deletion:		Insertion and Fissure:		
stage	stepwise	overall	stepwise	overall	
1	a	a	∅	∅	
	/\	│			
2	b c	│	a		
	│	│	/\		
3	b	b	b c	b c	(fissures obscured)

Furthermore, just as deletion and insertion obscure other operations on the deleted or inserted elements in (16), so also fusion obscures previous operations on the fused entities, and operations on entities coming from a fissure operation obscure the nature of the fissure. These are all the possible cases where an entity placed by one operation is itself the subject of another operation: each time, the steps are obscured in the overall mapping.

3.2.5 Representative Derivational History Relations

It is, then, in the nature of mapping over a series of steps that some operations may not be apparent in the final derivational history relation. It is useful to define a property of representativity that excludes this possibility:

(17) **Definition: Representativity of Correspondence under Serial Composition**

a. For sets A,B,C, and relations $R:A\rightarrow B$, and $S:B\rightarrow C$, the serial composition $SoR =_{def} \{\langle x,z\rangle \mid \exists y$ such that $\langle x,y\rangle\in R, \langle y,z\rangle\in S\}$ is **representative of S and R under serial composition** if and only if all and only those disparities in R or S are also disparities in the S o R mapping.

The serial composition of two mappings is representative if and only if all the disparities in the two indiviual mappings are visible in the overall mapping.

b. We shall also require the representativity property to be cumulative, so that given a further relation T:C→D, the serial composition T o (S o R) is **representative of T, S and R under serial composition** if and only if (i) it is representative of T and SoR under serial composition and (ii) SoR is itself representative of S and R under serial composition.

The serial composition of several mappings is representative if and only if all the disparities in each of the individual mappings are visible in the overall mapping.

This definition means that for a derivation of *n* steps, where ${}_{1}step_{2}$, ${}_{2}step_{3}$, ..., ${}_{n-1}step_{n}$ are the mappings at each step, the overall derivational history relation (18) is representative of the mappings at all the steps if and only if it contains the disparities that are present in all the mappings, ${}_{1}step_{2}$, ${}_{2}step_{3}$, ..., ${}_{n-1}step_{n}$.[2]

(18) Overall Representative Derivational History Relation

$${}_{1}H_{n} = {}_{n-1}step_{n} \text{ o } ... \text{ o } {}_{2}step_{3} \text{ o } {}_{1}step_{2}$$

Which derivations produce representative derivational history relations? Not those where an element produced by one operation is altered by another operation, for that leads to obscuring of the operations. Rather, derivations are representative when the rule operations in them alter elements of the structure in *different* parts of the structure (or more precisely, elements with distinct derivational histories). So representativity is co-extensive with the class of derivations in

[2] Unlike the basic definition in (10) in which composition proceeds in a properly nested fashion, the order of composition becomes irrelevant under representativity, and (18) reflects this. That is, for representative serial compositions, the serial composition operation is *commutative*.

which no rule may alter the material introduced by the structural change of another rule. Derivations conforming to this restriction we may call **cumulative derivations**.

Cumulative derivations still include some subtle rule combinations. For example, a feature change followed by metathesis is admissible because one changes the identity of the feature itself while the other changes an ordering statement for the feature tier:

(19) $-F_i \; -F_j$
$\quad\;\; \times$
$\quad -F_j \; +F_i$

Note that a phoneme segment may be affected more than once in a cumulative derivation, provided each *feature* is affected only once, since these are the formal objects of the phonological representation to which the cumulative restriction applies. The alternants [d]-[s]-[ʒ] in *decide-decisive-decision*, *evade-evasive-evasion*, *corrode-corrosive-corrosion*, display processes of spirantisation and palatalisation of the underlying /d/, but voicing alternates also (Chomsky 1995:224, Chomsky and Halle 1968:229ff). /d/ spirantises and devoices before the suffix -*ive*, so we might seek to apply these processes again before -*ion*, and then voice the fricative again for *decision*, etc. by the /s/-voicing rule of English (Chomsky and Halle (1968:228), cf. *gymnasium* [dʒɪmneɪziəm]). However, a cumulative derivation that starts with [+voice] will either keep it or devoice, but not *re*voice. Chomsky and Halle's chosen analysis conforms to this, since they derive [ʒ] in *decision* from /d/ by spirantisation and palatalisation as in (20). Devoicing applies with the -*ive* suffix and with a few other derivatives of verbs: *save* → *safe*, *use* → *usage*, but not with -*ion* (Chomsky and Halle 1968:232).

(20) Abbreviated derivation of *decisive* and *decision*

dəsaɪd ɪv	dəsaɪd iən	
z	z	Spirantisation (stem-final alveolars)
	ʒ	Palatalisation (of CiV)
s		Devoicing (stem-final fricatives)

3.2.6 Summary

Building on the analogy in 3.1 between rule operations and Faithfulness constraint violations, we have shown that the relation of derivational history is, formally speaking, the derivational counterpart to input-output correspondence relations.

We have shown that the application of operations in series is more complex than the violations of Faithfulness constraints in input-output maps, because one rule operation can change an object created by another operation. This leads to a derivational history relation that is not representative of all the steps. By contrast, input-output correspondences are always representative of themselves – a degenerate case of representativity. In **cumulative** derivations, however, no operation changes an object created by a previous operation. The derivational history relation is representative of all the steps, and so these derivations are no more complex than input-output correspondences. An input-output correspondence relation with *n* Faithfulness violations correlates with the class of *n*! possible cumulative derivations containing the analogous *n* operations and no others.

We go on to consider a more restricted range of correspondences.

3.3 Natural Correspondences

Even when hidden operations are ruled out, there are many ways a given pair of underlying and surface structures might correspond. For example, a mapping from *ab* to *ba* could be construed as a total relation with a reversal of order of the elements, or it could be construed as an empty relation in which the *a* and *b* in the final structure bear no relation to the *a* and *b* in the original structure. Still other relations are logically possible. This raises the question as to whether we can pick out the most natural correspondences, and whether we can constrain generation systems and evaluation systems so that only the natural correspondences are obtained.

3.3.1 Veritable Correspondence

Consider again the example from Veneto Italian from 3.2.4. A vowel-raising-and-lowering derivation like prɛ́ti → príti → prɛ́ti is complex specifically in that it does not provide the most natural underlying-surface correspondence.

The raising of mid vowels in stress position (head vowel of the rightmost foot) is achieved by the leftward spreading of a [+high] feature. Now this is accompanied by the delinking of the [-high] feature originally associated to that vowel - that is, the deletion of the association between it and its anchor, as in (21):

(21)

In derivations of words with non-ATR vowels like prɛ́ti → príti → prɛ́ti, raising of stressed mid vowels would produce the vowel /ɪ/, which being unattested in the language, will be the subject of a repair which in this dialect would reverse both operations: the [+high] feature must be delinked again (the inserted association must be deleted) and also the [-high] feature must be re-linked to the head vowel (an association must be inserted), as shown in (22).[3]

(22)

The deletion of a previously inserted association line (the spreading and delinking of [+high]) was discussed in 3.2.4, but we also have *insertion* of material identical to that *previously deleted* (delinking and relinking of [-high]). Now despite the Duke-of-York character of this derivation, the derivational history relation is perfectly representative here: the underlying V~[-high] association and the surface V~[-high] association do not correspond to each other, as illustrated in (23a), where they are indexed differently (*2* and *4*), representing the fact that the grammar removes an association and adds an association.

(23)	a. Insertion-&-Deletion mapping	b. Natural mapping
Underlying	V_1 $|_2$ $[-high]_3$	V_1 $|_2$ $[-high]_3$
	↓	↓
Surface	V_1 $|_4$ $[-high]_3$	V_1 $|_2$ $[-high]_3$

[3] There is indeterminacy as to how the spreading/delinking/relinking/delinking derivation would actually work. Once the [+high] first spreads in (21), does the [-high] delink to remove the anomaly of the dual specification for height as occurs in the simpler case of ATR vowels as we have assumed, or does the [+high] now delink again without delinking the [-high]? Or do both delink simultaneously? Or does [-high] delink before all other operations? Or simultaneously with the original spread of [+high]? There is nothing to go on to decide.

The one in (23b) represents the situation in which nothing is changed between underlying and surface forms, this reflected in the coindexing of all parts (*1*,*2*,*3*), including the association line. The mapping in (23b) eliminates two gratuitous disparities at a stroke from the underlying-surface correspondence: the lack of a surface correspondent to the underlying association, and the lack of an underlying correspondent to the surface association. In Veneto Italian, for example, this is achieved by adopting an analysis in which the raising rule is formulated so as not to delink [-high] in the first place if the vowel is [-ATR].

Whereas (23a) is subtle and surprising, (23b) is natural and expected. Why? Because it expresses a visible similarity. To give it a name, it is a "**veritable**" correspondence, expressing the similarities that are apparent between the two sub-structures. A veritable correspondence has just those disparities necessary to express the discrepancies, and none that would fail to express the similarities: it has a *minimum* of disparities. A veritable correspondence is not the same as a faithful correspondence in optimality theory. The usefulness of the term *veritable* lies in the fact that two structures may not be identical - so that there is no fully faithful correspondence between them - but there will still be a veritable correspondence between them, the one with the minimum of disparities. This allows us to express *how similar* structures are. How this works out precisely we now examine in the context first of the optimality framework, then of the derivational framework.

3.3.2 Minimal Violation of Faithfulness Constraints

In optimality theory, the minimisation of disparities amounts to the minimisation of Faithfulness constraint violations. According to the core Optimality-Theoretic principle of Violability (McCarthy and Prince 1994:336), constraints are violable in optimality theory, but *minimally*: only in order that higher-ranked constraints may be satisfied (or at least violated less seriously themselves). In particular, then, minimal violation applies to the Faithfulness constraints. And if Faithfulness constraints are violated minimally, disparities between underlying and surface structures will be minimised, so that candidates with veritable correspondences will be prefered.

For example, for each underlying form there will be candidates that employ input-output correspondences with complementary MAX and DEP violations between objects that happen to be identical in value and position, but such candidates will always be less harmonic than those with the veritable correspondence which lacks these violations. (24) evaluates correspondences between identical entities in the input and output: in candidate *a.* the *x*'s are not in correspondence, in *b.* they are.

(24)

		/..x../	IDENT	LINEARITY	MAX	DEP
a.	Input	x				
	Output	x			*	*
☞b.	Input	x				
		|				
	Output	x				

There is no constraint for which (24a) is more harmonic than (24b) - the Faithfulness constraints unambiguously prefer (24b), and no output constraint discriminates between them. (24a) is always less harmonic than (24b) whatever the ranking of constraints. It is therefore unoptimisable in a system that evaluates correspondence relations by Faithfulness constraints[4], a fact we represent here by shading not only the crucial violation marks but the candidate itself.

[4]It would be optimisable in a system which includes "anti-Faithfulness" constraints that *require* disparities.

Minimising the lack of correspondents as in (24) is counterbalanced by the similar avoidance of gratuitous multiple correspondents. Multiple correspondents violate the Faithfulness constraints INTEGRITY and UNIFORMITY. Candidates with complementary violations of these constraints are always less harmonic than candidates without the offending multiple correspondents. Hence, candidates a. and b. in the illustrative tableau (25) can never be optimal.

(25)

		/ x y /	INTEGRITY	UNIFORMITY
a.	Input Output	x y | / | x y	*	*
b.	Input Output	x y |X| x y	**	**
c. ☞	Input Output	x y | | x y		

Thus, complementary violations of MAX and DEP or of INTERGRITY and UNIFORMITY may be sifted out, reducing the breaches of totality and surjectvity, and functionality and injectivity.

In the cases just reviewed, the choice is between violating Faithfulness and not violating Faithfulness. Minimising violations is relatively straightforward. When it comes to IDENTITY and LINEARITY constraints, two correspondences have a chance of being optimal depending on how constraints are ranked. To fully understand the nature of veritable mapping, we shall now be occupied with examining various alternative formal mappings. Consider, for example, mapping from *xy* to *yx*. This might be viewed as reordering, resulting in a LINEARITY violation, or as the loss of one item and appearance of an identical one at a different site, resulting in MAX and DEP violations. Either analysis is possible for the *xy*→*yx* mapping depending on the ranking of the constraints, as shown in tableaux (26) and (27). (Since the ranking of MAX and DEP is not crucial we assume MAX >> DEP without loss of generality.)

(26)

/ x y /	MAX	DEP	LIN
a. Input x y / Output y x	*!	*	
☞b. Input x y X Output y x			*

(27)

/ x y /	LIN	MAX	DEP
a. Input x y / Output y x		*	*
☞b. Input x y X Output y x	*!		

A conflict arises between the properties of order preservation on the one hand, and totality and surjectivity on the other, since not all can hold at once in these circumstances. are possible depending a conflict arises since a violation of either may be alleviated by twin MAX and DEP violations, depending on how they are ranked. Similarly with a map –F→+F: the identity property conflicts with totality and surjectivity: MAX_F,DEP_F >> IDENT_F leads to optimal $–F_i \rightarrow +F_i$ with an IDENT violation; IDENT_F >> MAX_F,DEP_F leads to optimal $–F_i \rightarrow +F_j$ with twin violations of MAX and DEP. So either formal conception of neutralisation of feature values is possible: a value switch, or omission and replacement of the underlying feature.

Nevertheless, in these cases, one of the two correspondences is still recognisable as the veritable one. Conservation of material while merely changing the order (LINEARITY violation) or value (IDENT violation) is simpler because there is just one disparity breaching just one property, whereas the other alternative (MAX violation and DEP violation) has two disparities breaching two properties.

The veritable correspondence may be assured by a stronger appeal to the tenet of minimal constraint violation, drawing on an insight due to Pulleyblank and Turkel (1997). Just as *within an individual grammar* minimal constraint violation determines the relative harmony of candidates, it also evaluates *between grammars*, since constraint violations incurred by a candidate may be less serious in one grammar than in another. Pulleyblank and Turkel (1997) explore the consequences of this for the learnability of constraint hierarchies. Here, if we apply this approach to Faithfulness constraints specifically, we can derive verity of correspondence.

Violations of MAX, DEP, and LINEARITY are more serious in some grammars than others, depending on the ranking. In (28) we look at the candidates from tableaux (26) and (27) over three grammars. Candidate a., which violates MAX and DEP, is most harmonic in the grammar where these are ranked lowest - that is, LIN>>MAX>>DEP. Candidate b., which violates LIN, is most harmonic in the grammar where LIN is ranked the lowest, MAX>>DEP>>LIN. The minimal possible violation is provided in the grammar MAX>>DEP>>LIN where candidate b. has a single violation of the lowest ranked constraint (LIN). Both the grammar and the candidate are awarded the pointing finger to identify this case. Every other case has violations of the first- or second-ranked constraints. Cells are shaded where candidates are ruled out not only in their own grammar, but across the grammars.

(28)

	LIN	MAX	DEP
a. $y_j x$		*	*
b. $y_j x_i$	*!		
	MAX	LIN	DEP
a. $y_j x$	*!		*
b. $y_j x_i$		*	
☞	MAX	DEP	LIN
a. $y_j x$	*!	*	
☞b. $y_j x_i$			*

There is potentially a problem of bootstrapping here, if we select the veritable correspondence of candidate b. by using candidate b. to select the grammar MAX>>DEP>>LIN. To avoid bootstrapping, we must begin from a position of agnosticism about the correspondence relation for the input-output pair 〈 *xy* , *yx* 〉. All possible correspondences between the pair provide possible candidates; we want the candidate violating constraints as minimally as possible by selecting the grammar in which the violated constraints are ranked as low as possible. By adopting this approach, the veritable correspondence emerges.

There is a further consequence concerning which correspondence emerges from applying minimality across grammars. Input-output correspondences with one, two, three, or any number of violations of LINEARITY will always be more harmonic than twin violations of MAX and DEP. Here, the use of ranking to demote violated constraints means that *a greater number of constraints violated implies a mapping is suboptimal even if the total number of violations is the same or less than another alternative*. This is illustrated this in tableaux (29-31):

(29)

	/ x y z /	MAX	DEP	LIN
a.	x y z \ \ z x y	*!	*	
b. ☞	x y z z x y			**

(30)

	/ w x y z /	MAX	DEP	LIN
a.	w x y z \ \ \ z w x y	*!	*	
b. ☞	w x y z z w x y			***

(31)

	/ x y z /	MAX	DEP	LIN
a.	x y z | z y x	*!*	**	
b.	x y z z y x	*!	*	*
c. ☞	x y z z y x			***

This notion of verity says that two strings that draw on an identical set of letters are more naturally said to be related by reordering than by omission and replacement, an intuitively acceptable result.

This covers the various possible interactions between the Faithfulness constraints. At this point, we have clarified the notion of veritable correspondence so that it conforms to the following definition:

(32) **Definition: Verity of Correspondence**

An underlying-surface correspondence is **veritable** if it breaches the smallest possible number of the natural properties of relations.

The natural properties of relations were given in (1), and the Faithfulness constraints correspond to these properties. Minimality of constraint violation ensures that as few Faithfulness constraints as possible will incur violations, and, *at least up to this point*, this leads to veritable correspondences as defined in (32). We test this one step further in the next section.

3.3.3 Insertion, Deletion, and Relocation

The relative naturalness of reordering or insertion/deletion has bearing on the analysis of patterns in which items delete and insert in complementary contexts, such as the r~zero alternation word-finally in Eastern Massachusetts English (McCarthy 1991, 1993, 1999c, Halle and Idsardi 1997) and some other dialects. Consider an utterance (36) in which *r* is lost from some positions and gained in others:[5]

[5] We employ the naïve assumption, purely for the sake of argument, that words are underlyingly *r*-less or *r*-final in accordance with how they are rendered orthographically. In fact, "there is often little basis for maintaining the etymological distinction… certainly language learners receive no evidence from which they could infer different underlying representations for *tuna* and *tuner*… More importantly, the facts about the **distribution** of *r* are robust

(33) “*What is your*[∅] *prefered reading*, *Derrida*[ɹ] *or*[∅] *Homer*[∅]*?*”

Whereas in other utterances, intrusive *r*s would be inserted, e.g. *Derrida*[ɹ] *and Hegel*, considerations of verity would require an analysis of (33) in which an intrusive *r* such as that in the intervocalic environment *Derrida_ or* would not be inserted if it could be supplied from another position from which *r* is being lost anyway, as in (34):

(34) /jɔɹ$_1$/ /dɛɹɪdæ/ /ɔɹ$_2$/ /hoʊməɹ$_3$/

“*What is your*[∅] *prefered reading*, *Derrida*[ɹ$_2$] *or*[∅] *Homer*[∅]*?*”

Similarly, the expression *"algebra**r** is difficult"* is liable to take an intrusive *r* (Halle and Idsardi 1997:332), but as successive identical consonants, the two *r*s may be difficult to articulate, leading to a dissimilation effect (35) in some dialects/speakers (Johansson 1973). On a veritable analysis, whichever of the two positions *r* survives in, it will be the progeny of the underlying *r*.

(35) /ældʒəbɹ$_1$ə$_2$/ /ɪz/

[ældʒəbɹ$_1$ə$_2$ɪz] or [ældʒəbə$_2$ɹ$_1$ɪz] “*Algebra is…/ Algeb**ar** is…*”

Such a relocation analysis conforms to the verity property, so is formally simple in that respect at least. Relocation is found in language games (Pig Latin ***l**atin* → *atin**l**ey*, Chomsky and Halle 1968:343) and more chaotically in speech errors such as spoonerisms (***s**ork **p**ausages* for ***p**ork*

---

and productive… speakers of this dialect regularly delete and insert *r*s in the expected places when confronted with neologisms, loan words, and the like.” (McCarthy 1999c:2)

***s**ausages*), but evidence from the core of natural language is lacking. Let us then consider whether minimality of constraint violation would force a relocation analysis.

Ranking arguments from minimality of constraint violation initially appear to require relocation in such a case. LINEARITY would need to be ranked low enough to prompt reordering under the right conditions. LINEARITY is not ranked below MAX, for if it were, relocation would be the strategy of choice for *all* unacceptable *r*s, leading to pronunciations like *[hoʊmɹə] for *Homer* and *[jəɹɔ] for *your*. Instead, we require that omission is the strategy of choice, and relocation a secondary strategy when there is a place for the *r* to go. This would follow from the ranking DEP >> LIN >> MAX ("do not insert unless relocation is impossible; relocate if omission would lead to insertion being necessary; omit unless relocation is possible"). DEP >> LIN >> MAX can be selected over LIN >> DEP >> MAX on grounds of less serious violations as in (36).

(36)

/ x y z /			
	LIN	DEP	MAX
a. x y z \ \ z x y		*	*
b. x y z / \ \ z x y	*!		
☞	DEP	LIN	MAX
a. x y z \ \ z x y	*!		*
b. x y z / \ \ ☞ z x y		*	

This is the situation with regard to alternative interpretations of a *given* input-output mapping. However, when we look at the system *as a whole*, there is a counter-argument against DEP >> LIN >> MAX. In forms where objects are inserted, the top-ranked constraint (DEP) is violated. Under the rival ranking LIN >> DEP >> MAX, which does everything by deletion and insertion and never allows reordering, the top-ranked constraint (LIN) is never violated, even though the second and third constraints down are sometimes violated together (as in candidate a. in (36)). There is a difference between the two arguments. Even though the outputs are the same and not in dispute, the ranking argument for DEP >> LIN >> MAX in (36) is based on a *conflict* between two alternative candidate analyses (the correspondence relation is part of the candidate – as shown in 2.3), whereas the counter-argument is based on applying the metric systematically so that all constraints that are violated at all in the language are ranked as low as technically possible. So whether or not verity of correspondence holds absolutely turns on whether we rank only to select between alternative correspondences in a given *case*, or as a general strategy of minimising violations or seriousness of violations in a given *grammar*. The more principled choice is the general one - in which case relocation *is* excluded and verity fails in general. This would save us from the possible difficulties in processing utterances in which relocation could occur over any distance, e.g. from one end of an utterance to another, and is consistent with the absence of relocation from normal language.

We have now demonstrated that in Optimality Theory, Verity of Correspondence derives from ranking arguments that decide between alternative analyses of the correspondence mapping on the basis of minimality of constraint violation. However, we have also demonstrated that the full application of this metric - 'rank-all-violated-constraints-low' - deviates from the verity property in the case of relocation over an arbitrary distance. In the process, we have shown that the central principle of minimal violation in Optimality Theory determines not only grammatical surface forms, lexical forms by Lexicon Optimisation (Prince and Smolensky 1993), and constraint rankings (Pulleyblank and Turkel 1997), but also correspondence relations. We now turn to the derivational framework.

3.3.4 Economy of Derivation

Our goal is to consider how disparities may be minimised order to guarantee natural correspondences between underlying and surface structures. In the derivational framework, the minimisation of disparities is achieved by the minimisation of rule operations. Generation systems in general offer any combination of operations in series and thus inevitably raise the prospect of unnatural derivational history relations that exceed Verity as well as Representativity. The canonical (but not the only) examples of this are the Duke of York gambits. For example:

(37)	Insertion-Deletion	$\varnothing \rightarrow x \rightarrow \varnothing$	exceeds Representativity
	Deletion-Insertion	$x_i \rightarrow \varnothing \rightarrow x_j$	exceeds Verity

Although insertion-deletion exceeds one property and deletion-insertion the other, they may both be excluded by the single requirement that derivations contain only the minimum of operations. The criterion of minimal derivation length has previously been used to select in syntax between two derivations with equivalent results (Chomsky 1995), to constrain the form of phonological rules (Calabrese 1995), and to limit the abstractness of underlying phonological forms (Kiparsky 1982).

A strict economy of derivation condition, slightly modified, will have as its consequence both the Representativity and the Verity of underlying-surface correspondences. Taking the question of verity first, the minimisation of operations will provide for the minimisation of disparities in the mapping between underlying and surface structures.

A surface form that is identical to the underlying form will be required to eschew deletion and insertion, fissure and fusion, double order change, and double value change, which all exceed verity, since the most economical derivation contains precisely no operations. Where a grammar contains two rules that are capable of reversing each other, the condition of economy will select rules that have sufficient restrictions on their application as to prevent creating a long derivation, as Calabrese (1995:447) has observed.

The derivation *xy*→*yx* will be done by the single operation of order change, avoiding longer possibilities such as deletion and insertion at different sites: $xy_i \rightarrow y_i \rightarrow y_i x$.

Multiple order changes become uneconomical where three or more are necessary, given the alternative of deletion and insertion at the different sites (38a,.b.). General "movement" operations would perform relocation in one fell swoop providing a yet more economical derivation (38c.), although a movement operation describes behaviour only seen in language games, so is presumably unavailable in normal language phonology.[6]

(38)

a.		b.		c.	
$wxyz_i$		$wxyz_i$		$wxyz_i$	
wxz_iy	Order change	wxy	Deletion	z_iwxy	Movement
wz_ixy	Order change	z_jwxy	Insertion		
z_iwxy	Order change				

---

[6] Non-adjacent metathesis *xyz*→*zyx* is also considerably less cumbersome when it is assumed that general movement operations are available in language play, but deletion-insertion may be worse than multiple order change:

xyz		xyz		x_iyz_j	
xzy	Order change	xy	Deletion	z_jyx_i	Movement
zxy	Order change	xyx	Insertion		
zyx	Order change	yx	Deletion		
		zyx	Insertion		

In the previous section, we claimed that these mappings make best sense as a re-ordering of the same segments – i.e. non-order-preserving. Economy of Derivation (39a.) fails to select the veritable correspondence in this instance because it minimises operations rather than types of operations (Deletion, Insertion, Fusion, Fissure, Value Change, Order Change). However, if operations were minimised by type (39b.), the multiple order changes would be selected, since only one type of operation is involved, keeping the verity of derivational mappings.

(39)

a. **Strict Economy of Derivation (standard)**

Adopt the grammar which produces derivations containing the fewest operations.

b. **Strict Economy of Derivation by Type of Operation**

Adopt the grammar which produces derivations containing the fewest types of operations, and then the fewest operations of each type.

Economy of Derivation by Type of Operation would favour relocation by a rule that applies precisely when one *r* is needing to be removed while an *r* is needed at another site, as in (34) and (35), since relocation reduces the types of operation used in the derivation from two (insertion and deletion) to one (change in order). This keeps verity but complicates the rule system with a conspiratorial metathesis rule. Standard Economy of Derivation, on the other hand, would not allow the unlimited order changes necessary to relocate *r*s to new intrusive contexts, since two operations, a deletion and an insertion, are sufficient. Of course, standard Economy of Derivation fails to respect verity in cases of multiple reordering generally, so relocation over an arbitrary distance is merely a specific case. Nevertheless, standard Economy of Derivation succeeds in avoiding the addition of an extra metathesis rule when insertion and deletion rules already produce the correct results, and it also avoids the potential high processing complexity of relocating segments to different parts of the utterance.

3.3.5 Economy and Representativity

We have shown how Verity is secured by Economy of Derivation by Type of Operation. We now show that Economy of Derivation by Type of Operation also ensures *Representativity* of Correspondence – which was defined and discussed in 3.2.5.

The nonrepresentative Duke of York gambits such as insertion-deletion ∅→x→∅ leave no trace whatsoever in the underlying-surface correspondence. Economy of Derivation (either formulation) excludes them. Then there are pairs of operations whose overall effect may be replaced by just one operation. Thus, the result of inserting a feature, ∅→-F, and then changing its value, -F→+F, may be obtained by just one operation, ∅→+F. Further checking shows that most other nonrepresentative combinations may similarly be replaced by one simple operation, but two cases remain: the first is where a fused element is subsequently deleted (40a), and the second is where an inserted element is made the subject of a fissure (40b).

(40) a.

	stepwise		overall
1	a b		a b
	\/		
2	c	Fusion	
3	∅	Deletion	∅

b.

	stepwise		overall
1	∅		∅
2	a	Insertion	
	/\		
3	b c	Fissure	b c

As is clear from (40a), the representative alternative to fusion-and-deletion is the deletion of both adjacent objects in the original structure. But this alternative also consists of two operations, no more economical than before. Similarly for (40b), the alternative to the Trojan Horse strategy of insertion off an object followed by breaking it into two is merely the insertion of two objects. Again, this still uses two operations, again no more economical. Only when we appeal to Economy of Derivation by Type of Operation can we discriminate against these in favour of the representative alternatives. Then deletion only is selected over fusion and deletion, and insertion only is selected over insertion and fissure.

Now, these scenarios are clearly unusual: (40a) represents a case where two objects are deleted in a given context, while (40b) represents a case where two inserted objects appear in a given context. For example, Wheeler and Touretzky (1993:170) note that a pair of insertion rules $\varnothing \rightarrow a$, $\varnothing \rightarrow b$ in the same context *p__q* are unattested in any one language – suggesting a fortuitous limit on computational complexity in language, which apparently need not contend with the problem of determining the relative linear order in which the objects *a*,*b* are to appear. Nevertheless, some objects may be inserted adjacently. Stem augmentation in Lardil nominatives, for example, employs some consonant-vowel sequences, e.g. *kan**ta*** vs. nonfuture accusative *kan-in*. Strict Economy of Derivation by Type of Operation requires in any such case that objects are inserted or deleted without the use of fusion or fissure. It seems unlikely that one would find empirical grounds for an insertion-and-fissure Trojan Horse gambit in a language (vowel epenthesis followed by diphthongisation of the vowel, for example, goes against the expectation that epenthetic vowels take the unmarked, default quality), but it is a formal possibility given the standard Economy of Derivation principle.

We have now seen how Strict Economy of Derivation by Type of Operation (35b) provides for both Representativity and Verity. So with one single condition, derivations may be confined to those whose mapping observes two complementary simple properties: excluding the introduction both of disparities that would *not* be reflected in underlying-surface correspondences, and of disparities that *could* be reflected there but would offer greater complexity than necessary.

3.3.6 Summary: Minimality of Violation & Economy of Derivation

Having identified two simple properties – representativity and verity – that might be applied to correspondence mappings between underlying and surface structures, we have found that these are achieved in the two frameworks if the familiar strategies of economy of derivation and minimality of constraint violation are modified slightly from their simplest form.

This is possible because the strategies of economy of derivation and minimality of Faithfulness violations have essentially the same form: as few Faithfulness constraints as possible are violated, as few times as possible, or, as few types of operations as possible are used, as few times as possible. The similar effects obtained by this common strategy of *minimisation* serve to further substantiate the basic analogy between operations and Faithfulness violations on which the formal comparison is built. The similar results are summarised in (41) below.

(41) **Results**

Derivational Framework:	**Verity**	**Representativity**
strict Economy of Derivation *by type of operation*	✓	✓
strict Economy of Derivation *standard version*	✗ Disfavours reordering more than twice	✗ Admits Trojan Horse derivations

Optimality Framework:	Verity	Representativity
Minimal Violation metric *applied in case of conflict*	✓	✓ (vacuously)
Minimal Violation metric *applied over language as a whole*	✗ Disfavours reordering over an arbitrary distance	✓ (vacuously)

We have shown in detail how Optimality Theory and Minimalist derivational theory make very similar predictions as to the complexity of the underlying-surface correspondence. However, they can be distinguished as to their explanatory value. Minimality of constraint violation lies at the very heart of Optimality Theory, being the very basis of optimisation, but in derivational theory, the Economy of Derivation principle must be specially added. Furthermore, in Optimality Theory, minimality of constraint violation determines four interlocking aspects of the phonological grammar – surface forms, underlying forms, constraint rankings, and underlying-surface correspondences. In derivational theory, however, the economy of derivation principle determines just two of these - underlying-surface correspondences (which we have demonstrated) and choice of underlying form (as discussed by Kiparsky 1982). If naturalness of correspondence can be sustained empirically – and in the next section we claim that it can – then it is Optimality Theory that offers greater explanatory depth here, since it explains the particular predictive properties of the special principle of Economy of Derivation within a more general scope.

3.4 The Duke of York Gambit

Having compared how the underlying-surface mapping is handled in the two theoretical frameworks, it remains to examine the viability of maintaining a natural correspondence between underlying and surface structures. In particular, if the “Duke of York gambit” proved to be necessary in phonology, as Pullum (1976) claimed, it would undermine both Optimality Theory and Minimalist derivational theory. We claim here that the unnaturalness of the Duke of York mapping not only explains phonologists' negative reactions to it, but also shows it to be unexplanatory. We then show that its exclusion is supported by the empirical evidence after all.

3.4.1 The Trouble With The Duke of York Gambit

Natural correspondences are often assumed implicitly in the literature with little or no discussion. In Optimality Theory, McCarthy and Prince (1995) pointed out that all manner of correspondence relations are possible for any given output, but in practice, only one intuitively obvious correspondence relation (the natural one) for a given output is at all interesting. Optimality theorists want to explain why this output is optimal and not that one; the question of why it is this correspondence relation and not that one is merely a matter of technical hygiene whose outcome is nearly always intuitively obvious. Similarly, rule analysts may well assume that the operations that apply are those which are evident in the resulting forms. Concerning rules of Lengthening and Shortening of stressed vowels in English, Halle (1995:27,28) observes that

"On the one hand, in words such as *divin-ity, natur-al* and *ton-ic, athlet-ic,* as well as in *Palestin-ian* the stem vowel is shortened, but there is no shortening in *ton-al* or *atone-ment*. On the other hand, there is lengthening in *Caucas-ian, remedi-al,* but not in *remedy-ing, buri-al* or *Casp-ian*."

"strings having the same form as the output of the more restrictive <Lengthening> rule are prohibited from undergoing the less restrictive <Shortening> rule… as a consequence, neither *Shakespear-ian* nor *jov-ial* are subject to Shortening".

Shortening applies to the head vowel of a branching foot, while Lengthening applies to the head vowel of a branching foot **provided** the vowel is [-high], the following vowel is /i/, and the vowel following that is in hiatus with /i/. What is most interesting is what happens when both rules could potentially be relevant in the same derivation. Halle's explanation as to why the head vowels in words such as *Shakespearian* and *jovial* are long is that Shortening is blocked. If this is so, then the rules apply only when it is clear from the surface forms that they have. In fact, as Prince (1997b) points out, it would be unnecessary to prevent Shortening from applying to *Shakespear-ian* and *jov-ial* since in these cases Shortening would be followed by Lengthening, still giving the correct result. The two options are shown in (42), where length is indicated by moras ($_{\mu}$ - short, $_{\mu\mu}$ -long):

(42)	jo$_{\mu\mu}$vial	or	jo$_{\mu\mu}$vial	
	blocked		jo$_{\mu}$vial	Shortening
			jo$_{\mu\mu}$vial	Lengthening
	[dʒoʊviəl]		[dʒoʊviəl]	

The second alternative would explain why the surface vowel is long without the need for an extra constraint on the application of Shortening. Only if it is implicitly assumed that underlying-surface correspondences are natural is it necessary to explain the long surface vowels by the nonapplication of Shortening.

An intuitive respect for the naturalness of the underlying-surface mapping by phonologists explains their long-standing though somewhat inchoate suspicion of the Duke of York gambit. Pullum (1976:84-85) cites a number of authors' comments on such analyses: "suspicious", "juggling", "complicate the description", "illegitimate practice", "unattractive", “farcical”, “missing something”. More recent authors continue the trend, calling the strategy “dubious” (McCarthy 1993a:182), an “implausible expedient” (McCarthy 1993b:7) and “undesirable” (Halle and Idsardi 1997:346). When Iverson (1989) offers an analysis of alternations of laterals in Klamath in terms of gemination and degemination, an apology is deemed necessary for the two-part creation and dissolution of a geminate in derivations /ll̥/→ l̥ː → [lh], /llˀ/ → lˀː → [lʔ], though that apology is no more than the offer of a mirage lacking in any factual support:

> "independence of its parts *might* lie in stylistic variation between the geminated, presumably careful speech product of [the first process] and reduced, possibly more casual output of [the second process], *although Barker (1964) himself gives no evidence of the distinction*" (Iverson 1989:297 *my italics*).

It is not the mutually-contrary *rules themselves* that are seen as risible: rather, even when an analyst is content to defend two contrary rules in a description, it is their occurrence in the same *derivation* that is considered curious. Thus, McCarthy (1991) demonstrates that both insertion and deletion operations are necessary in the analysis of stem-final *r* in Eastern Massachusetts English, but he rejects the possibility of one altering the outcome of the other in a derivation:

> This analysis may be a descriptive success, but it is an explanatory failure. The derivations are dubious, because many r’s are inserted at Word level only to be deleted phrasally in what Pullum 1976 calls the "Duke of York gambit". (McCarthy 1993a:182)

It is not made clear what such derivations supposedly fail to explain, but the author seems to imply that objects that have to be deleted should not have been inserted in the first place. The reference to Pullum does not help because his position was that “the Duke of York gambit will be reasonable precisely when the result is a reasonable analysis, and unreasonable precisely when it is not" (Pullum 1976:100). Nevertheless, Halle and Idsardi (1997:343ff) do precisely the same as McCarthy, arguing that *r*-insertion and *r*-deletion rules are both necessary for Eastern Massachusetts English, yet still insisting that the analysis must specifically avoid the two applying in the same derivation, even though the correct results can still be obtained that way.

These negative reactions to the Duke of York gambit, that have been persistent over the history of Generative Phonology despite the descriptive correctness of the analyses, may collectively be explained as the intuitive rejection of derivations that bring about a relationship between underlying and surface structures which is not the natural one.

3.4.2 The Duke of York Gambit as Unexplanatory

The Duke of York gambit fails to explain why the surface form is so similar to the underlying form – and, therefore, to the other alternants on the basis of which the underlying form is set up. On the Duke-of-York analysis, they are similar by accident, because a rule happens to re-create the same configuration as was there to start with. The similarity of the surface form to its underlying form is only *explained* by a grammar that *preserves the configuration as it is*.[7] Duke of York derivations create a derivational history which exceeds the naturalness that is necessary to explain similarities in form between alternants; properly economical derivations explain similarities in alternants. Indeed, the empirical record shows that formal naturalness of the underlying/surface relation is preserved, so that phonology explains both why forms are as they are and why they are similar to their alternants. Other uneconomical derivations are unexplanatory. Metathesis $x_1y_2 \rightarrow y_2x_1$ is superior to uneconomical deletion and re-insertion at the new site because deletion and re-insertion fail to explain why the string has

[7]The similarity between the forms may be explained historically, if one of the rules in the Duke of York analysis arose as a hypercorrection of the other rule: Halle and Idsardi (1997) claim that Eastern Massachusetts English *r*-insertion arose as a hypercorrective rule, though they still prefer to block the Duke of York derivation in the synchronic grammar of current Eastern Massachusetts English.

identical elements, if in a different order. The unnatural relocation *over an arbitrarily large distance*, $x_i.....\varnothing \rightarrow \varnothing......x_i$ (see 3.3.3), however, would be an over-ambitious attempt to "explain" the identity of objects that are being systematically added and removed from unconnected sites, whereas the derivationally economical analysis using only insertions and deletions is also a more straightforward description of effects at unrelated sites.

This account of the explanatory inadequacy of the Duke of York gambit locates it in a broader issue affecting unnatural underlying/surface relationships, and replaces previous, unviable criticisms of the Duke of York gambit, such as the objection of Halle and Idsardi (1997:344):

> *r*-deletion preceding *r*-insertion results in derivations where the effects of *r*-deletion are repaired by *r*-insertion. This type of interaction has been termed by Pullum 'the Duke of York gambit' and objections to it have been raised on the grounds that the gambit subverts the essential difference between rules, which reflect idiosyncratic facts of a language, and repairs, which are consequences of general structural principles obeyed by the language.

Objecting that the gambit subverts the difference between rules and repairs assumes that the second rule is somehow 'acting as' a repair. This is merely in the eye of the beholder. Repairs are context-free rules that change structural configurations; they are unordered and apply persistently (Myers 1991). In the analysis discussed here, *r*-insertion is clearly not a repair rule but an idiosyncratic, ordered rule (Halle and Idsardi 1997:343-4). The *r*-deletion rule removes *r* from codas, and the *r*-insertion rule is subject to an array of conditions that would not be placed on a repair: a hiatus-filler in coda (rather than onset) position; only following non-high vowels; and only in derived environments (or rather at Prosodic Word boundaries - McCarthy 1999c:8-9). The choice of *r* as the inserted segmental structure is a particular and complex phonological stipulation (McCarthy 1993a, Blevins 1997), not a simple structural adjustment that would be characteristic of a repair. It is simply incorrect to say that *r*-insertion **repairs** the overapplication of *r*-deletion: it just **reverses** it. It is indicative of the very broad, formal nature of the Duke of

York gambit that a rule may conceivably be reversed by an unordered, persistent, repair rule or by a subsequently-ordered idiosyncratic rule.[8]

What the Pullum (1976) study showed was that the Duke of York gambit is "an independent issue, cutting across many other issues in phonological description" (Pullum 1976:94). It is independent of opacity, abstractness, and stipulative rule ordering (Pullum 1976:89-94), but also independent of the ordered rule / repair rule distinction. It has to do with constructing an underlying-surface correspondence, and is not the most simple way of doing so.

3.4.3 Eastern Masshachusetts r

It is true that some Duke-of-York derivations have arisen in eccentric analyses which are disputable on independent grounds (Pullum 1976:88). But even the case of Eastern Massachusetts English *r*, which is amenable to a Duke of York analysis at least for the most familiar data in (43), does not go through when additional data is considered. The basic pattern of *r*~∅ alternation is as follows:

(43)

a. Before a vowel	*I put the tuna on the table*	*I put the tuner on the table*
	[…tjunəɹ ɒn…]	[…tjunəɹ ɒn…]
b. Not before a cons.	*I put the tuna down*	*I put the tuner down*
	[…tjunə daʊn]	[…tjunə daʊn]

[8]If one repair A→B was accompanied by another B→A (a curious theoretical possibility noted by Scobbie 1991:18), then each would repeatedly reverse the other in a derivation that fails to terminate. Repetition is disallowed by a universal constraint banning reapplication of either unordered rule in interrupted sequence (Ringen 1976). However, according to Myers (1991), repairs apply "persistently", in direct contradiction to Ringen, in which case mutually-reversing repairs *would* fail to terminate.

c. Not before a pause *I found the tuna* *I found the tuner*

[…tjunə] […tjunə]

d. Before class II affix

drawing; withdrawal *layering; empowering*

[dɹɔɹiŋ] [wɪðdɹɔɹəl] [leɪəɹɪŋ] [ɪmpaʊəɹɪŋ]

The pattern, also present in other dialects including British English and Australian English, has attracted analyses using insertion only (Johansson 1973, Pullum 1976), deletion only (Gick 1999), or both (McCarthy 1991,1999c, Halle and Idsardi 1997) (further references given in McMahon, Foulkes and Tollfree 1994). Evidence for an underlying distinction between *r*-final stems (subject to deletion) and *r*-less stems (subject to insertion) comes in a few items from their realisation with class I affixes such as -*ic* (*Homer*/*Homeric*, *alter*/*alteration* vs *Volta_*/*Voltaic*, *algebra_*/*algebraic*), and "speakers regularly delete and insert *r*s in the expected places when confronted with neologisms, loan words, and the like" (McCarthy 1999c:2). This much demonstrates that both deletion and insertion of *r* are positively motivated. The contexts "before a consonant" and "before a pause" indicate that *r* is deleted in coda position, but the more open, lenited, and vocalic quality of inserted *r* demonstrates that it is inserted in coda position also. So if *r*s are deleted from the coda, but re-inserted into coda before a vowel to account for (43a), then a Duke of York derivation obtains, and this correctly derives the data in (43).

However, the data in (44) will not be derived correctly. The *r*s on the right are lenited (marked as less constricted), matching the characteristic phonetics of coda *r*.

(44) Procliticised Function Words (McCarthy 1999c)

r Not Inserted:	***r* Present:**
to Ed	*for Ed*
[tə ɛd]	[fəɹ̣ ɛd]
to add to his troubles	*for any reason*
[tə æd tə ɪz tɹəbəlz]	[fəɹ̣ ɛnɪ ɹizən]
why do Albert and you	*they're eating*
[wɑj də ælbət ən juw]	[ðeəɹ̣ ijɾɪŋ]

Insertion must be restricted so as not to apply to words on the left, by restricting it to apply at the right boundary of a phonological word and not at a proclitic-word juncture (McCarthy 1999c:9). But then the words on the right would be subject to coda *r*-deletion in a context where insertion does not restore the *r*, wrongly leaving them *r*-less. Coda *r*-deletion must therefore be restricted so as not to apply to the coda *r*s on the right. One possibility might be that *r*-deletion, like *r*-insertion, applies only at the right boundary of a phonological word - but this misses the fact that the absence of coda *r* is very general except for a few special positions (word-finally and proclitic-finally). A more satisfactory alternative is to re-formulate the generalisation so that *r* is deleted from *all strict* codas - not from prevocalic codas, which are *ambisyllabic* (Kahn 1976) and therefore the *r* is licensed by its onset affiliation even though it inherits the phonetic quality of the coda. And *this* precludes a Duke-of-York interaction between *r*-deletion and *r*-insertion, since it means that prevocalic *r* in *"Homer is difficult"* will not now be deleted and in need of re-insertion.

3.4.4 Evidence Against The Duke of York Gambit

Pullum (1976:100) was led to conclude that Duke of York derivations are an inevitable and integral component of the description of at least a residue of cases of mutually contrary processes in a derivational system, and CANNOT justifiably be disallowed (see also Zwicky 1974, Prince 1997b). This lines up with the dictum of Kiparsky (1982:172): "derivational simplicity is strictly subordinated to grammatical simplicity, and only comes into play when the evaluation measure is indeterminate as between alternative grammars." On this view, preference for shorter derivations would be relative, not absolute. In fact, the evidence - hitherto unrecognised - goes against Pullum's conclusion. Consider a language with both epenthesis and syncope of vowels. Their combined application in derivations would be detectable because the whole vowel inventory would collapse in that context down to the epenthetic vowel quality:

(45) $\{i,e,a,o,u\}\rightarrow\varnothing\rightarrow\{i\}$

Alternatively, failure of combined application would leave the vowel inventory intact. Not all syncopating languages are testing grounds for inventory collapse vs. preservation: there may be one vowel which undergoes both syncope and epenthesis (e.g. /i/ in Palestinian Arabic, Brame 1974), in other languages, vowels of any quality may undergo syncope but there is no epenthesis reported (Maltese, Brame 1974; Southeastern Tepehuan, Kager 1997). But general vowel syncope and *i*-epenthesis come together in Yokuts (Kuroda 1967, Kisseberth 1970a). In Yokuts, collapsing of the vowel inventory does not occur; rather, the two processes occur precisely in *complementary* contexts as in (46):

(46)	**Epenthesis**: CC_C or C_C]		
	xat-t	[xat*i*t]	‘ate’
	paʔṭ-hin	[paʔ*i*ṭhin]	‘fight’-aorist
	Syncope: VC__CV		
	kili:y-*a*-ni	[kileyni]	‘cloud’-indirect objective
	xat-a-ni	[xata:ni ~ xatni]	‘eating(N)’-indirect objective
(cf.	polm-a-ni	[polma:ni]	‘husband’-indirect objective)
	hall-(h)at*i*n-i:n	[hallatnen]	‘lift up’-desiderative-future
(cf.	bint-(h)atin-xu:-k’a	[bintatinxok’]	‘be trying to ask’)[9]

Cross-linguistically, syncope consistently fails to apply just when it would create consonant clusters that do not fit the syllable structure of a language (Myers 1991:318), while vowel epenthesis is repeatedly employed to break up consonant clusters that do not fit the syllable structure. So syncope and epenthesis occur in disjoint contexts systematically. It is also the case that “it is not unusual to find processes of vowel deletion creating complex clusters that are broken up by epenthesis in a different place, such as /yiktibu/→*yiktbu*→*yikitbu* in some dialects of Arabic.” (McCarthy 2002:169-170). An Amerindian language, Chukchee, displays vowel syncope and epenthesis at different sites in this way:

[9]The complementary nature of syncope and epenthesis means that instances of /i/ may either be analysed as syncopated in the contexts where they do not appear, or epenthesised in the contexts where they do appear. Thus Archangeli (1985:347) analyses the desiderative suffix as /atn/, with epenthesis possible between the /t/ and /n/. However, this will not work for instances of syncopated /a/, demonstrating that a syncope process is operative in Yokuts in addition to an epenthesis process.

(47) (Kenstowicz 1994:106)

	abs.pl.	erg.	abs.sg.
‘load’	imti-t	imti-te	imət (imti→imt→imət)
‘box’	cenle-t	cenle-te	cenəl (cenle→cenl→cenəl)
‘walrus fat’	lonla-t	lonla-te	lonəl (lonla→lonl→lonəl)

By contrast, Duke of York interactions between syncope and epenthesis applying at the *same* site are to my knowledge unattested.

In another example, loss of the final stem vowel is used morphologically to mark the nominative for Lardil nominals (Prince and Smolensky 1993:97ff, Hale 1973), yet short words are phonologically augmented by word-final epenthesis. Again, although a Duke-of-York derivation would be possible in certain words, it does not in fact occur. In (48a), apocope-marked nominatives are shown, where vowel absence is denoted by "_"; in (48b), monomoraic stems (those containing just one short vowel) are shown augmented in the nominative by an epenthetic **a** vowel; in (48c), bimoraic stems (containing precisely two short vowels), which might have been susceptible to both subtraction and augmentation, in fact exhibit neither.

(48)	Nominative	Nonfuture Accusative	Future Accusative	Gloss
a. **3+ Vowels: Subtraction**				
/yiliyili/	yiliyil_	yiliyilin	yilyiliwuŗ	’oyster sp’
/mayařa/	mayař_	mayařan	mayařaŗ	’rainbow’
b. **1 vowel: Augmentation**				
/yak/	yak**a**	yakin	yakuŗ	’fish’
/ŗelk/	ŗelk**a**	ŗelkin	ŗelkuŗ	’head’
c. **2 vowels: Neither**				
/mela/	mela	melan	melaŗ	’sea’
/wiţe/	wiţe	wiţen	wiţeŗ	’inside’

A stem like *mela* does not settle the matter, since it may or may not derive by *a→∅→a*. However, the preservation of the stem *wiţe* in the nominative shows that there is no *e→∅→a* derivation. Impoverishment of {i,e,u,a} to {a} does not take place in the final vowels of bimoraic nominatives, providing decisive evidence that subtractive nominative marking and word augmentation apply in complementary contexts in Lardil, not at the same site.[10]

The overall pattern is that impoverishment of the vowel inventory in some context down to a single vowel, the epenthetic vowel, is not found; instead, there are cases where vowel

[10] Our proposal for disjunctivity goes against a hypothetical reconstruction by Blevins (1997) of the historical development of truncation in Lardil. As well as final vowel deletion there is a process that lowers final vowels. This predates the vowel deletion process, when it would have applied to all stems, e.g. */yalulu/ →*[yalula], but is now only visible on the bimoraic stems which resist deletion, e.g. /ngawu/→[ngawa]. Blevins (1997:255) claims that "subsequent to V-lowering.., the final [a] of the stem could be reanalysed as an instance of nominative /-a/, with deletion of preceding stem-final /u/. Subsequently, the paradigm was levelled by loss of all nominative /-a/s, in favour of the ∅ form of the case-marker, still occurring in some /i/-final and all /a/-final stems like /mela/ ’sea’." The

deletion and insertion apply in disjoint contexts, and cases where vowels are inserted at a different site to the deletion site. This holds true both for syllable-based patterns of syncope and epenthesis (Yokuts, Chukchee), and for word-based patterns of subtraction and augmentation (Lardil). It remains to be seen whether a case of impoverishment by contrary processes might be brought to light, but on the available evidence we claim that this is not possible.

3.4.5 Resyllabification and Trans-Domain Interactions

Another possible type of evidence for a Duke of York derivation would be where a rule crucially depends on the intermediate stage between two contrary rules and so necessarily intervenes between them. If it necessarily intervenes between them, then clearly they must both apply - before and after the intervening one. McCarthy (2003) has reanalysed some putative examples of this and claimed that such cases are generally lacking, but there remain some special subcases.

For example, a segmental process may be conditioned by a syllable structure which is subsequently erased and replaced. In such examples, the two contrary operations are ones which build and erase constituent structure (e.g. syllable structure). Thus, syllabification rules may feed certain segmental processes dependent on syllable structure, followed by the erasure of that syllable structure. In an example from Attic Greek (Noyer 1997:510), adjacent nuclear vowels are subject to a rule of Contraction which brings them together as one syllable, e.g. p^h*i.(lé.e).te* → p^h*i.(lée).te.*. The second of the two syllables is built and then destroyed, though its earlier presence was crucial to the analysis in ensuring correct placement of antepenultimate stress opaquely on *.lé.*, not on *.*p^h*í.* as expected from the surface form. The special feasibility of resyllabification effects is amenable to explanation using Optimality Theory: see McCarthy

---

middle stage of her reconstruction contains precisely the scenario we have excluded - truncation (e.g. /u/-loss) and augmentation (/-a/) together.

(2003) for an account using sympathy theory, and chapter 6 for an account in a theory of serial cumulation of constraints.

In addition to this, McCarthy (2003) notes data from Hebrew which shows that Duke of York gambits may occur due to rules with different domains of application. In words that bear the prepositional prefixes /bi#/ 'in' or /ki#/ 'like', postvocalic spirantisation (a process discussed again in 6.2.2) may be crucially conditioned by a vowel that is inserted and subsequently deleted again. Thus, in one stratum, the stem is subject to epenthesis and spirantisation [bi[ktob]] → [bi[kətob]] → [bi[kəθoβ]], operations also used to produce the free-standing word kə.θoːβ 'writing'. In a subsequent stratum, other rules including removal of schwa apply to the domain that includes /bi#/ deriving the final form bikəθoβ → bix.θoːβ . The sequence of mappings across the two strata contains the ∅→ə, ə→∅ combination so is not representative under serial composition, but within each stratum the mapping is representative. Alternatively, in Optimality Theory, this can be interpreted as a paradigm uniformity effect between the bi# form and the free-standing form (McCarthy 2003). The free-standing form exhibits vowel insertion and postvocalic spirantisation, and the spirant is carried over to the bi# form by paradigm uniformity even in the absence of a conditioning vowel there.

The Duke of York gambit *can* have explanatory value if the intermediate stage is crucially necessary to condition another process. Then the Duke of York gambit would crucially explain why the output attested that process. This applies to the cases attested here. Another uneconomical derivation that can be explanatory is feature insertion followed by value change ∅ → [-F] → [+F], which has explanatory value if the feature actually behaves as [-F] for purposes of some process even though the surface value is [+F]. We shall argue in 6.2.5 that cases of this kind also occur.

3.4.6 Summary: A Universal Property

The evidence, not only from the lack of intervening rules (McCarthy 2003), but also from the maintenance of vowel inventories in the face of potential collapse in languages with vowel deletion and vowel insertion, suggests that a disjunctive interaction between contrary processes is a universal property of language – the opposite of Pullum's conclusion (Pullum 1976:100):

(49) **Universal Interaction of Mutually Contrary Operations**[11]

Mutually contrary operations on segmental structure apply in distinct contexts.

This would explain the phenomenon of the maintenance of vowel inventories as an inevitable part of language, that does not need to be specially learned – learners do not have to find a way to block vowel deletion to avoid collapsing the inventory. It can be derived either from Economy of Derivation or from Minimal Constraint Violation. It could be proved false if any evidence comes to light of prevarication on the part of learners or the adopters of fresh contrary processes, or of a stable impoverishment pattern in any language.

The universality of the interaction is expected to apply to morphological operations, as in Lardil, as well as phonology proper. In Optimality Theory, the formalism appears at first sight to allow for the possibility that (49) fails specifically when it comes to morphological operations. In Lardil there is a conflict between the phonology and the nominative marking which is resolved by satisfying the FOOT-BINARITY constraint "Feet $(...)_{Ft}$ contain two moras" at the expense of violating the REALISEMORPHEME constraint which requires that morphological categories (here, nominative case) be marked by some deviation from the plain stem (Kurisu 2001). The tableau in (50) illustrates this:

[11] The notion of 'mutually contrary' processes will be given a precise treatment in 4.3.

(50)

/wiţe$_i$/	FT-BIN	DEP(V)	REALISE MORPHEME	MAX(V)
☞a.(wiţe$_i$)$_{Ft}$			*	
b. (wiţ)$_{Ft}$	*!			*
c. (wiţa$_j$)$_{Ft}$		*!		*

If DEP(V) were not crucially ranked above REALISEMORPHEME, candidate c. would be optimal. This predicts that the stem-final vowel would be replaced by *a* in bimoraic stems, apparently simulating an uneconomical *e*→∅→*a* derivational sequence. However, the input/output analysis ⟨ *wiţe$_i$* ,*wiţa$_j$* ⟩ is an unnatural correspondence - as Prince and Smolensky (1993:112) intuitively recognise when they call it a "devious analysis". Therefore, it would be ruled out if the Minimal Violation Metric is given full application (see 3.3.2), avoiding twin MAX and DEP violations. Hence, a genuine instance of a morphological deletion that is reversed by phonological insertion would constitute a counterexample to the minimal-violation-metric approach.

3.4.7 An Unsuccessful Proposal: The Elsewhere Condition

Halle and Idsardi (1997) claim that Duke of York derivations are precluded by the Elsewhere Condition. Although it is claimed that the Elsewhere Condition (Kiparsky 1973) is “among the most important contributions to phonology” (Halle and Idsardi 1997:344) and “an empirical result of some importance” (Halle 1995:27), no evidence is offered for the impossibility of combined application of any pair of mutually contrary rules with properly nested environments, nor *for* the necessity of combined application of any pair of mutually contrary rules whose environments are *not* properly nested. Both pieces of evidence are required if the proposal is to be tenable, but as we have seen, such evidence has to be subtle because the two derivational strategies usually provide equally successful descriptions. As pointed out by Prince (1997b), this leaves no apparent reason to augment rule theory with the special condition.

It is true that in the original paper on the Elsewhere Condition, Kiparsky (1973:100ff) claims to show that in Vedic Sanskrit, two rules in a special/general relationship implicating the values [+syllabic] and [-syllabic] on high vowels, are found not to apply in the same derivation. He predicts that the intermediate stages of putative $w \rightarrow u \rightarrow w$ / $j \rightarrow i \rightarrow j$ derivations would be expected to be retained in certain metrical contexts, and yet are not. Even if Kiparsky is right about Vedic Sanskrit (Howard 1975 disputes this), the result is merely consistent with our general claim that Duke of York interactions do not occur, and does not necessarily support the view that only those in a special/general relationship are blocked.[12] McCarthy (1999c) shows that the Elsewhere Condition cannot account for all the data in the case of Eastern Massachusetts English *r* (Halle and Idsardi 1997).

3.5 Conclusion

There is a substantial analogy between operations of rules and violations of Faithfulness constraints that provides a productive area of formal comparison. The correlation is stronger when derivational history relations are representative of their derivational steps, a property which does not hold of all derivations.

The Economy of Derivation Principle of Minimalism, and the Minimal Violation Metric of Optimality Theory, confine grammars to natural correspondences which exclude various

---

[12] In fact, of the two rules, one changes [-syllabic] to [+syllabic] and one changes [+syllabic] to [-syllabic], which means that the structures affected by one rule are not nested in those affected by the other rule. They only appear to be in a special/general relationship when they are expressed in an implicational format, with the feature [syllabic] mentioned only in the outcome and not in the structural description of the rule. This format is questionable because, although such rules achieve greater simplicity, implicational rules have to be interpreted differently from the usual transformational rules (see 2.1.2).

intuitively unexpected maps such as deletion and re-insertion at the same site, or relocation of a segment over an arbitrary distance. Optimality Theory offers greater explanatory depth, making similar predictions to those of the special principle of Economy of Derivation within its own more general scope.

The elimination of the unnatural Duke-of-York mappings is correct on empirical grounds, given evidence from patterns of deletion and insertion of vowels. The Duke-of-York proscription includes morphology as well as phonology proper. The proscription also allows us to explain similarities between alternant forms, rather than leaving these similarities as accidental consequences of certain rules. Recalcitrant Duke-of-York gambits are confined to special subcases: the building and erasing of constituent structure leading to ‘resyllabification’ effects, and cases that may be interpreted as paradigm uniformity effects, as in Hebrew (see 5.3 for another type). In chapter six, we will provide a theoretical system whose predictions and limitations prove to coincide with these empirical observations.