6:

CONSTRAINT CUMULATION THEORY

This chapter presents a new formal synthesis of derivational theory and optimality theory, **Constraint Cumulation Theory** (CCT). Phonological constraints are arranged in an order which serves both as a serial order and a rank order: earlier constraints are **pre-eminent** over later ones in that they may take effect first, then persist alongside later ones, remaining dominant over them. Phonological representations of utterances, then, are derived by constraints which *accumulate* one by one.

Our first task will be to present the system and demonstrate its descriptive coverage of the interactive effects among phonological processes recognised in chapter four: *overapplication*, *mutual interdependence*, and *default*. We also demonstrate its selective coverage of Duke-of-York interactions, which accords with the findings of chapter five.

Then, we further examine patterns of underapplication and overapplication, ultimately predicting from the theory that neutralisation and deletion processes will remain transparent except under certain conditions, whereas insertion and conditioned specification of features are expected to become opaque. This successfully resurrects a long-abandoned problem (Kenstowicz and Kisseberth 1973, 1977).

6.1 Towards Multifarious Interaction

An adequate theory of phonological grammar must be able to accommodate all the various interactions between generalisations witnessed in the phonology of the world's languages. This includes the ability to accommodate all the examples in (1).

(1)

a. Supporting Generalisation: Vowel Raising in non-standard Italian (Calabrese 1995)

/prɛt i/ → [priti] (raising of vowel supports tensing of high vowels)

*[prɪti] (raised, but untensed)

b. Overapplication: Epenthesis in High-Vowel/Liquid Sequences in Eastern Massachusetts English (McCarthy 1991)

/fijɹ/ → [fi.jə] (epenthesis overapplies with respect to loss of *r*)

*[fi.jəɹ] (epenthesis, but no *r*-deletion)

*[fij] (*r*-deletion removing the need for epenthesis)

c. Mutual Interdependence: Stem-Syllable Alignment & Coda Licensing in Lardil (Prince and Smolensky 1993)

/kaŋ/ → [kaŋ.ka] (stem ends at syllable edge with nasal coda licensed)

*[ka.ŋa] (stem and syllable misaligned)

*[kaŋ.a] (nasal coda unlicensed)

d. Default: Vowel Syncope in Yokuts (Kisseberth 1970a)

/kiliːy a ni/ → [ki.ley.ni] (one vowel omitted, CV(C) syllables maintained)

*[ki.leyn] (vowel omissions, unacceptable coda cluster)

*[ki.le.yin] (vowel omissions compensated by *i*-epenthesis)

As was demonstrated formally in chapter 4, rule-based generation systems and constraint-evaluation systems are only partially successful at capturing this range of patterns. Their performance is summarised in table (2) below. In its standard conception at least, Optimality Theory cannot handle (b) *overapplication*. On the other hand, rules cannot handle (c) *mutual interdependence* at all; (d) *default* is expressible only rather awkwardly by imposing an additional "blocking" subsystem on the core rule system (see 2.2.2, 4.2.2).

(2)

	Rule-based Derivational Theory	Optimality Theory
a. Supporting Generalisation	✓	✓
b. Overapplication	✓	✗
c. Mutual Interdependence	✗	✓
d. Default	✗	✓

However, by combining the strengths of both derivational theory and optimality theory in a cumulative system of constraints, all these patterns can be described.[1]

6.1.1 The Constraint Cumulation System

We propose that a phonology consists of a set of constraints on representations **M1, M2, M3, ..., Mn** ('M' for Markedness broadly construed), and that these are interspersed with Faithfulness constraints *F1, F2, F3, ..., Fm* in an ordered list - e.g. *F1,* **M1,** *F2,* **M2, M3,** *F3,* **...,** *Fm*, **...,** **Mn**. The constaints accumulate in a series of steps:

[1]Declarative phonology (Scobbie, Coleman and Bird 1997) does not accommodate (b) or (d), although extended versions have accommodated (d) *default* constraints.

(3)

- Stage 1: if an underlying representation does not satisfy **M1**, then the grammar derives from it the representation that best satisfies the hierarchy

  F1 >> **M1**>> *F2* >> *F3* >> ... >> *Fm*

- Stage 2: if that representation does not satisfy **M2**, then the grammar derives from it the representation that best satisfies the hierarchy

  F1 >> **M1**>> *F2* >> **M2**>> *F3* >> ... >> *Fm*

- *etc.*

Or, more generally:

- Stage *k*, for *k*=1,2,3,...,*n*: if the representation at stage *k* does not satisfy **Mk**, then the grammar derives the representation that best satisfies the hierarchy containing **M1, M2, ..., Mk** (but not Mk+1,...,Mn *k*<*n*) interleaved with all of the Faithfulness constraints *F1, F2, ..., Fm* .

According to this approach, a phonological grammar has constraints placed in **cumulative order**. Each Markedness constraint is added in turn, so at successive stages an increasing number of constraints are operative. There is *serial* interaction because the constraints are added one at a time, there is *hierarchical* interaction because each constraint participates at all subsequent steps as a higher-ranked constraint, and there is *mutual* interaction because at each step previous constraints appear together with the newly added constraint. Where one constraint is ordered earlier than another, we shall say that that constraint is **pre-eminent** over the other, this pre-eminence having both serial and hierarchical implications. The name Constraint Cumulation Theory reflects the fact that the various interactional effects witnessed in phonology emerge as constraints accumulate. So, later constraints may render opaque the effect of earlier constraints; later constraints may act in concert with former constraints; and later constraints will be of null effect when in conflict with earlier constraints. The effects of the constraints on the representation at each stage are determined in conjunction with the Faithfulness constraints, which are *all* present at every stage even though they are interspersed among the Markedness constraints in the ordering.

Faced with the choice, *constraints* rather than rules are adopted as the elements of phonological grammar. The difficulty with rules is that each Structural Description is fused with its own particular Structural Change, which does not allow at all for outcomes created by mutually supporting generalisations such as (1c). This requires several generalisations at once to influence a single step, as discussed by Prince and Smolensky (1993:124-125). Other difficulties with rules are pointed out in Scobbie (1991), Singh (1987), Sommerstein (1974), Shibitani (1973) and Kisseberth (1970a). In abandoning rules, CCT adopts the constraint-based view that processes are enacted to avert the violation of constraints. However, building derivations from the interaction of Markedness and Faithfulness constraints improves on earlier approaches that used constraints in derivations by providing not only for serial interaction, but also for the default behaviour and mutual interdependence between phonological generalisations highlighted in the Optimality Theory literature.

6.1.2 Descriptive Coverage

We illustrate CCT's descriptive coverage of the examples from non-standard Italian, Eastern Massachusetts English, Lardil and Yokuts that were given in (1). In the Italian pattern of vowel metaphony, vowel raising and vowel tensing both occur in forms like [priti] *priests*, from /prɛt/. We recognise the following two constraints as responsible for these processes:

(4) ALIGN = ALIGNLEFT([+high],Foot)

Every feature [+high] must be aligned to the left of a foot.[2]

*+high/-ATR

Vowels cannot have both [+high] and [-ATR].

We illustrate the cumulation of these constraints in the order ALIGN before *+high/-ATR. As laid out in (5), raising would be brought about at the step at which the Alignment constraint is included, and tensing would be brought about at the next step where *+high/-ATR is included. This, then, works as a "feeding" order: raising is achieved at one step, tensing at the next step.

(5)

/prɛti/	ALIGN
prɛti	*!
☞prɪti	

↓

/prɪti/	ALIGN	*+high/-ATR
prɛti	*!	
prɪti		*!
☞priti		
preti	*!	

↓

/priti/

[2]According to Calabrese (1995:445), metaphony is restricted to *final* feet. As it stands, the Alignment constraint does not incorporate this restriction, and so is correct only for words with one foot.

The same result obtains if ALIGN and *+high/-ATR accumulate in the opposite order. In that case, *+high/-ATR is included first, but to no effect, since the form /prɛti/ already adheres to the constraint. At the subsequent step when the Alignment constraint is also included, the mutual effect of the two constraints will be to force the result /priti/ at a stroke, as shown in (6).

(6) /prɛti/

/prɛti/	*+high/-ATR
☞prɛti	
prɪti	*!

↓

/prɛti/	*+high/-ATR	ALIGN
prɛti		*!
prɪti	*!	
☞priti		
preti		*!

↓

/priti/

A supporting generalisation, then, will take effect irrespective of the cumulative order. The two kinds of cumulation illustrated will each capture other important effects: the step-by-step nature of cumulation allows for the possibility of overapplication of some processes, while the presence of many accumulated constraints within a given step allows for the possibility of mutually interdependent processes. We illustrate these next.

The example of overapplication we have selected is the ban on tautosyllabic sequences of high vowel and liquid in Eastern Massachusetts English. In words like [fijəl] *feel*, epenthesis breaks up the offending sequence, but in words like [fijə] *fear*, epenthesis and *r*-deletion both occur, and the latter process removes one of the two segments that conditioned the epenthesis process. The relevant constraints are in (7).

(7) *[+hi][+ap] $]_\sigma$ = *[+high][+consonantal,+approximant] $]_\sigma$

No tautosyllabic sequences of high vowel and liquid.

*Coda/r

No *r* in the coda.

Cumulative Order:

*[+high][+consonantal,+approximant] $]_\sigma$ before *Coda/r before MAX-C

Placing *[+hi][+ap] $]_\sigma$ earlier in the order guarantees that it will take effect even though it is not surface apparent, in the classic serial-derivational fashion.

(8) /fiːr/

	*[+hi][+ap] $]_\sigma$	MAX-C
fiːr	*!	
☞fi.jəɹ		
fiː		*!

↓

/fijəɹ/	*[+hi][+ap] $]_{\sigma}$	*Coda/r	MAX-C
fiːr	*!	*	
fi.jəɹ		*!	
☞fi.jə			*

↓

/fijə/

Notice the importance of Faithfulness constraints participating at each step. At the second of the two steps, the optimal form violates MAX-C, and since this violation is not fatal MAX-C must be ranked lower than the two markedness constraints. Nevertheless, it is crucial that Faithfulness constraints are present at all steps, so that MAX-C is present at the first step to rule out /fi:/. Were /fi:/ to be selected, it would stand as the final outcome of the grammar - contrary to fact - since /fi:/ satisfies both constraints just using the single process of *r*-deletion, pre-empting epenthesis. This is precisely what we want to avoid if overapplication is to take effect.

Having shown that supporting-generalisation effects and overapplication effects can be readily accounted for in CCT, we may recall that these effects even occur simultaneously, as was demonstrated in 4.2. One such case is Turkish epenthesis and *k*-deletion (Orgun and Sprouse 1999): in derivations of words like inek-m → inek-i-m → ine-i-m ‘my cow’ with 1sg possessive suffix *-m*, epenthesis supports the occurrence of intervocalic *k*-deletion by supplying one of the vowels, while *k*-deletion removes one of the consonants in the cluster that prompted vowel epenthesis, causing it to overapply. Such a case is straightforward in CCT. For present purposes, we may employ an *ad hoc* constraint *Vk]V that simply rules out *k* in the context where it is not found - intervocalically, in stem-final position. If we place this constraint within a cumulative set, NoComplexCoda before *Vk]V before Max-V before Max-C before Dep-V, then the result falls out correctly:

(9) /inekm/

/inekm/	NoComplex Coda	Max-V	Max-C	Dep-V
inekm	*!			
☞inekim				*
inem			*!	

↓

/inekim/	NoComplex Coda	*Vk]V	Max-V	Max-C	Dep-V
inekim		*!			
inekm	*!				
☞ineim				*	
inem			*!	*	

↓

/ineim/

Having demonstrated supporting-generalisation effects (1a) and overapplication effects (1b), Lardil provides our example of mutually interdependent processes (1c). If a noun stem is short, the nominative form will consist of the stem plus epenthetic segments to bring it up to disyllabic length. In the example [kaŋ.ka], augmented from stem /kaŋ/, an epenthetic *k* provides a syllable onset, with stem *ŋ* placed in the coda. The constraints in (10) are jointly responsible for this outcome.

(10) CODACOND

A coda has no Place specification of its own, unless it is the unmarked Coronal place.

ALIGNRIGHT(Stem, Syllable)

The right edge of a stem coincides with the end of a syllable.

Since these constraints act in a mutually interdependent fashion within a single step once they are both present, it is irrelevant which comes first in the cumulative order. We adopt the order CODACOND *before* ALIGNRIGHT(Stem, Syllable).[3] This yields the derivation in (11).

(11)

/kaŋ/	CODACOND
kaŋ	*!
kaŋ.a	*!
☞ka.ŋa	

↓

/ka.ŋa/	CODACOND	ALIGNRIGHT	DEP-C
kaŋ.a	*!		
ka.ŋa		*!	
☞kaŋ.ka			*

↓

/kaŋ.ka/

[3]As it happens, there is other evidence from Lardil in support of this order. In [ya.ka], from /yak/, the restriction on codas is respected and the final underlying consonant is syllabified as an onset, at the expense of true stem/syllable alignment (see Prince and Smolensky 1993 for full discussion). Stem/syllable alignment must therefore come later in the order, its effects being blocked by earlier constraints in such cases.

The syllabification of *ŋ* in the onset at the first step counter to fact is easily revised at the second step, because there are no Faithfulness constraints requiring the retention of syllable structure (this is due to the fact that syllabification is not contrastive in language - see McCarthy 2003). Now at the second step, one or other of the two constraints can be satisfied depending on the syllabification of the *ŋ* as kaŋ.a or ka.ŋa. However, both constraints can be satisfied at once at this step if an output with several fresh characteristics is selected: *ŋ* syllabified in the coda, and an inserted homorganic stop syllabified in the following onset. Then the edge of the stem /kaŋ/ does coincide with the syllable boundary, yet the Coda Condition is also respected since the Dorsal place specification of the *ŋ* is not exclusive to the coda but is shared with the following onset.

Having seen how CCT predicts both overapplying processes and mutually interdependent processes, we finally turn to default processes. Syncope displays default behaviour cross-linguistically, since it fails to occur where it would leave behind consonant clusters that cannot be incorporated into other syllables. We assume that syncope is the loss of vowels falling outside of feet, following the analysis of Southeastern Tepehuan in Kager (1997). In (12) we give the relevant constraints for syncope in Yawelmani Yokuts.

(12) NOCOMPLEX

Syllable onsets and syllable codas must not contain two or more segments.

PARSE-σ

Syllables must belong to feet.

Cumulative Order:

NOCOMPLEX, MAX-C before DEP-V before PARSE-σ before MAX-V

Since syncope defers to overriding syllable structure constraints, they must come earlier in the cumulative order than the syllable-to-foot parsing constraint which minimises vowels outside feet. This leads to the derivation of /kiliːy a ni/ as in (13). There is an initial foot /ki.liː|/ (we shall not be concerned with the assignment of feet - see Archangeli 1991) and vowels outside it violate PARSE-σ. As a result, one is removed. We shall not be concerned here with the choice of vowel to be syncopated, but with the issue of how the constraints suppress excessive vowel removal and compensatory epenthesis, which deviate from the default behaviour that is observed.

(13) /ki.liː|yani/

	NOCOMPLEX	MAX-C	DEP-V	MAX-V
☞ki.liː|.ya.ni				

↓

/ki.liː|.ya.ni/

	NOCOMPLEX	MAX-C	DEP-V	PARSE-σ	MAX-V
ki.liː|.ya.ni				**!	
☞ki.liːy|.ni				*	*
ki.liːyn|	*!				*
ki.liː|.yin			*!	*	**

↓

/ki.liːy|.ni/

↓ (other changes)

/ki.ley|.ni/

In (13), there are potentially two unfooted syllables, and one is removed. Outputs with excessive vowel elision, deriving a *yn* coda or adopting a compensatory epenthetic ***i*** (/ki.liːyn|/, /ki.liː|.yin/), are rejected.

Neither would it be possible to delete a vowel at *one* step and insert ***i*** at a *subsequent* step, *yan→yn→yin*, because the NOCOMPLEX constraint that causes epenthesis is also present to prevent the *yn* cluster in the first place. Even if NOCOMPLEX were put *after* PARSE-σ in the cumulative order, so that vowel deletion *yan→yn* is enforced at the first step by PARSE-σ in the absence of NOCOMPLEX, it would still not be possible to epenthesise once NOCOMPLEX were added, since the more dominant PARSE-σ would block any epenthesis outside the foot. Neither cumulative order produces re-insertion. Thus, CCT correctly rules out collapse of a vowel inventory down to the default vowel by deletion-and-epenthesis, which we argued in chapter 3 does not occur in language.

6.1.3 Selective Admission of Duke of York Derivations

The absence of these deletion-and-(re)insertion derivations, which would be revealed by inventory collapse *V→∅→i*, demonstrated that the Duke-of-York derivations that are possible in rule-based phonology are not attested in language. We also cited McCarthy (2003) who claims that there is no evidence that intervening rules make crucial use of the intermediate stage between one structural change and its subsequent reversal. An example would be if syllables with (say) open vowels were lengthened to form heavy *Ca:* syllables, influencing rules of stress placement, then were made light (*Ca*) again in their surface form.

In CCT, these derivations are impossible. Cumulative order means that where two constraints make conflicting requirements on some context, the earlier one triggers its own process, and then blocks the outcome of the later one as appropriate. The later constraint is unable to alter any effect of the earlier one owing to the continuing dominance of the earlier one. So, as just seen, epenthesis is triggered by syllable structure restrictions, and the pre-eminence of these restrictions blocks the opposite process of syncope, despite pressure from the syllable-to-foot-parsing constraint to remove vowels; for another example, labialised velars [k^w] in Nootka are not permitted word-finally and are delabialised, but the constraint also blocks the opposite process of labialisation in the word-final position, even if the velar follows a rounded vowel, which is the right conditioning environment for labialisation: [...ok] *...*ok*^{w} (Pullum 1976, Sapir and Swadesh 1978); and so on. In all such cases, neither constraint *reverses* the effects of the other. Furthermore, since no process is applied and then reversed, there can never be any processes crucially triggered at some intervening stage - an apparently correct prediction.

Nevertheless, the two kinds of Duke of York derivation shown in (14) are attested, as pointed out in 3.4.5 and 5.3.3. These particular subtypes *can* be handled by CCT.

(14)

a. Reprosodification, e.g. Resyllabification: Compensatory Lengthening in Kinyarwanda (Walli-Sagey 1986)

/kuŋgana/ → [kuː.ᵑga.na] (long vowel takes timing slot of nasal)

*[kuŋ.ga.na] (inadmissible CVC syllable)

*[ku.ᵑga.na] (no extra timing slot in 1st syllable)

b. Reversal in Absence of Conflict: Palatalisation before front vowels and Depalatalisaton before coronals in Slovak (Rubach 1993)

/let En aː/ → [letnaː] (alveolar before another coronal)

*[let´ɛnaː] (vowel realised; cons palatalised)

*[let´naː] (prepalatal consonant before coronal)

First, resyllabification. In Kinyarwanda, the timing sequence is set up by one syllabification, and preserved when syllabification changes. In the derivation in (16), the nasal contributes a timing unit to the first syllable at the first step. When it is ultimately realised as prenasalisation of the following stop, its timing slot is retained and allotted to the vowel, so that vowel length indirectly marks the presence of a nasal in the lexicon.

(15) PARSE-Segment

Segments must belong to syllables

NOCODA

Syllables do not contain codas.

MAX-μ

Moras must be preserved.

Cumulative Order:

PARSE-Segment before NOCODA before MAX-μ

These provide the following derivation:

(16) /kuŋgana/

	PARSE	MAX-μ
☞kuŋ.ga.na		

↓

/kuŋ.ga.na/		PARSE	NOCODA	MAX-μ
	kuŋ.ga.na		*!	
	ku.ᵑga.na			*!
	☞kuː.ᵑga.na			

↓

/kuː.ᵑga.na/

In fact, change to syllable structure commonly accompanies change to the segmental string at successive steps, as has already been witnessed in the examples from Eastern Massachusetts English, Lardil and Yawelmani Yokuts. Assignment of new syllable structure at each stage is possible in CCT because there are no Faithfulness constraints covering prosodic structures, whereas there *are* Faithfulness constraints requiring the maintenance of features, segments and timing units conditioned *by* earlier syllable structures. Confinement of Faithfulness constraints to feature, segment and timing structure correctly accounts for the fact that only this material is contrastive in phonology (McCarthy 2003), and also successfully predicts the existence of resyllabification effects once the proposal to accumulate constraints over successive steps is adopted.

The facts in Slovak represent a different situation, where two conditions - for and against palatalisation - are relevant to mutually exclusive contexts. There are prepalatal consonants in Slovak, but they do not occur before coronal consonants, where they will be depalatalised, while dento-alveolars do not occur before front vowels, where they will be palatalised.[4] There can be

[4]Palatalisation does not apply morpheme-internally, e.g. [teraz] 'now'. This suggests the existence of a constraint that preserves the status quo when the underlying root is not prepalatal. On the other hand, underlying /l´/ (but not /t´,d´,n´/) is resistant to Depalatalisation (Rubach 1993:268), so a further constraint preserves underlying lateral prepalatals. We shall not pursue these details of Slovak here, but we will consider further the case for constraints that preserve underlying specifications in 6.2.

no blocking of palatalisation by depalatalisation, or vice versa, simply because they never conflict. Instead, they interact indirectly as follows. In Slovak, palatalisation overapplies with respect to deletion of the yer vowel /E/ (yer vowels are lost except when preceding a syllable containing another yer vowel) as seen in the alternation [strela] 'shot'(N), [strel´ba] /strel+Eb+a/ 'shooting' (Rubach 1993:114); Depalatalisation "undoes the effect of Coronal Palatalization whenever Coronal Palatalization has applied before a yer-initial suffix containing a coronal consonant and the yer has not been vocalized" (Rubach 1993:267).

For the three conditions involved, we adopt the cumulative order *No Palatal (__Coronal)* before *Palatal (__Front V)* before *No Vocalisation* (i.e., "yer deletion"). This order takes into account the fact that palatalisation overapplies with respect to yer deletion. In (17), we see the steps of the derivation as the requirements of Palatal and No Vocalisation are added.

(17) /let En aː/

	No Palatal (__Cor)	Palatal (_E)
letEnaː		*!
☞let´Enaː		

↓

/let´Enaː/

	No Palatal (__Cor)	Palatal (_E)	No Vocalisation
let´Enaː			*!
letEnaː		*!	*
☞letnaː			
let´naː	*!		

↓

/letnaː/

The coronal /t/ becomes prepalatal temporarily in the derivation, palatalising and depalatalising as its context changes. Since No Palatal does not conflict with Palatal, it does not prevent temporary palatalisation, even if it is pre-eminent.

This completes a survey of CCT's descriptive coverage of the different kinds of interaction between phonological generalisations. CCT, which was designed for the purpose of combining serial, hierarchical and mutual interactions of constraints, has the additional, unintended consequence that it admits the attested Duke-of-York derivations of resyllabification and reversal in absence of conflict, but does not admit unattested Duke-of-York derivations.

6.2 Predicting Underapplication and Overapplication

We have already seen how overapplication works in Constraint Cumulation Theory. Now, we observe that, although CCT is designed to incorporate serial interaction, the translation from rule-based derivations into this system is not complete. The interaction between Markedness and Faithfulness constraints in a cumulative system may be used to describe various underapplication phenomena, but it is sufficiently different from rule interaction to derive interesting constraints on overapplication and underapplication.

6.2.1 The Failure to Reproduce "Counterfeeding"

Constraint Cumulation Theory cannot directly reproduce the counterfeeding interaction that produces *underapplication* effects among rules, and instead will depend on pre-eminent Faithfulness constraints to cause underapplication. Whereas an ordered rule will fail to apply to the outputs of later rules, a constraint in a cumulative ordered set will instead be present as a potential filter at all subsequent stages. To see this, consider an example of a putative counterfeeding analysis from Malay. Low vowels in unstressed, light syllables reduce to schwa.

(18) Malay final vowel reduction (Zaharani 1998:144)

[dúgə][dugáʔan]	'to test, testing'
[súkə] [sukáʔan]	'to like, favourite'
[mṹlə][mũláʔi]	'to begin, to cause to begin'
[tʃíntə][tʃintáʔi]	'to love, to cause to love'

Standard usage in Malay omits word-final *r*, although the current prescription in Malaysia based on literary Malay requires the *r* to be pronounced (Zaharani 1998:87). This creates examples of low vowels in unstressed, open syllables in the following words.

(19) (cited by Zaharani 1998:145)

bakar	[baka]	'burn'
tukar	[tuka]	'change'
kisar	[kisa]	'revolve'
kindar	[kinda]	'avoid'

The low vowels in the words in (19) are NOT reduced to schwa. This might suggest a counterfeeding rule-based analysis in which the *r*-deletion rule is ordered after the low-vowel-reduction rule so that the output of *r*-deletion does not undergo vowel reduction. Such an analysis cannot be reproduced in terms of cumulative constraints. Recasting the structural descriptions of the rules as constraints, we have a constraint prohibiting final light syllables with [a] ("NOLIGHT[a]" in Zaharani 1998) and another prohibiting syllable-final [r] ("ALIGN-RHOTIC" in Zaharani 1998). We do not obtain the same results if we reproduce the rule order "Vowel Reduction before *r*-Deletion" with NOLIGHT[a] before ALIGN-RHOTIC in a cumulative system. For the stems in (18), NOLIGHT[a] will correctly force vowel reduction, but vowel reduction will be triggered incorrectly on the stems in (19). The NOLIGHT[a] constraint is added to the active hierarchy first but to no effect; once the ALIGN-RHOTIC constraint is added, NOLIGHT[a] and

ALIGN-RHOTIC are both active, rejecting both *r*-final forms like [kisar] and *a*-final forms like [kisa]. The system incorrectly selects schwa-final forms like [kisə] to be the grammatical forms, since these satisfy both constraints.

In the Malay case, the counterfeeding analysis has in fact been discredited on grounds that the data in (19) is phonetically inaccurate. Whenever *r*s are deleted word-finally, the adjacent vowels are in fact compensatingly long, as in the data in (20). The extent to which vowel reduction occurs in Malay is accounted for merely if it is confined to short vowels. Indeed, the confinement of alternations to short elements and the corresponding stability of long elements is already recognised in phonology as the general principle of *geminate inalterability* (Steriade and Schein 1986).

(20) Malay final compensatory lengthening (Zaharani 1998:87,88)

[kotoː]	[kotorran]	'dirty, dirt'
[ukeː]	[ukerran]	'to carve, carving'
[ukoː]	[ukorran]	'to measure, measurement'
[pasaː]	[pasarran]	'to market, market'

The failure to reproduce the counterfeeding interaction for the flawed Malay example in CCT reveals a general characteristic of the structure of cumulative constraint systems as divergent from ordered rule systems. Whereas a rule will not apply if it is ordered earlier, a cumulative constraint that is ordered earlier persists in its effects at later stages - it is even dominant over later constraints. This we take to be a potentially desirable result, for suspicion has consistently fallen on the counterfeeding analysis in phonology. It has been seen as a last resort, since it accounts for underapplication by a stipulation that is both language-specific and formally abstract (Booij 1997:277). Authors have sought to explain the failure of processes in other, more general ways, such as structural principles (like geminate inalterability in Malay) or affiliation of the process to a stratum of rules applying only in the context of certain affixes. A strong proscription on counterfeeding *commits* the theorist to finding principled explanations of why processes fail. We claim that the failure of processes is always due to some pre-eminent constraint, whether it be a Markedness constraint or a Faithfulness constraint:

(21) **Why Processes Fail:** due to pre-eminent constraints

a. Blocking using a Markedness constraint (*the process is excluded wherever it would create a proscribed phonological structure*)

b. Prespecification using a Faithfulness constraint (*the process goes against an overriding requirement of faithfulness to a specification at the previous stage, so the process is only possible where no specification is provided*)

Blocking of syncope was witnessed in 6.1, and the Malay example is also a case of blocking. Failure of processes due to prespecification is an approach promoted by Inkelas, Orgun and Zoll (1997). We shall extend it here in the context of the Constraint Cumulation system, applying it in the next section to underapplication effects in Modern Hebrew.

6.2.2 Prespecification: Modern Hebrew

Idsardi (1997) claims that in Modern Hebrew, postvocalic spirantisation overapplies with respect to syncope, and underapplies with respect to various processes of degemination, merger and deletion - purportedly supporting an ordered-rule analysis in which counterfeeding effects occur. Constraint Cumulation Theory can satisfactorily account for the overapplication of spirantisation in relation to syncope by placing a constraint against postvocalic stops into a cumulatively-ordered constraint set, and it can simultaneously account for the default behaviour of syncope by the pre-eminence of constraints on syllable structure, but the apparent underapplication of spirantisation in the data cited by Idsardi (1997) is best understood in terms of faithfulness to specifications at previous stages of the grammar, extending the prespecification approach of Inkelas, Orgun and Zoll (1997).

The process of postvocalic spirantisation in Modern Hebrew (MH) is retained from Tiberian Hebrew (TH), the language of the Bible as annotated by the Masoretes (Idsardi 1997, Bolozky 1978).

(22) Modern Hebrew spirantisation (Idsardi 1997)

i.	pakax	’opened eyes’	jifkax	’will open eyes’
ii.	lispor	’to count’	bisfor	’on counting’
	liʃpox	’to pour’	biʃfox	’on pouring’
	limkor	’to sell’	bimxor	’on selling’

In (22i), we see a [p~f] alternation in the first stem consonant depending on whether it is preceded by a vowel. In (22ii), the second consonant of the stem alternates, despite the absence of preceding vowels in the surface forms. In the forms on the right, spirantisation is triggered by a preceding vowel which is syncopated. In the forms on the left, syncope operates in the domain [*li* + stem] as part of the morpho- or lexical phonology, so that spirantisation, a word-domain process, operates on CVCCVC forms and so fails to spirantise the third consonant. In Tiberian Hebrew, stop-spirant variation was fully predictable - so the distribution of [+continuant] and [-continuant] is entirely specified by the grammar, and may be left unspecified in the lexicon (Idsardi 1998, 1997:372). Modern Hebrew has the spirantisation pattern, with a couple of special characteristics: only labials and voiceless velars [p~f b~v k~x] alternate in MH, and non-alternating /p b k f v x/ stops and fricatives occur as well as alternating ones. (23) lists various examples of non-alternating /p b k f v x/.

(23) Modern Hebrew /p b k f v x/ (quoted in Idsardi 1997)

a. *Historical merger* /k/ < TH /q/

kara	jikra	'read, will read'
kana	jikne	'bought, will buy'
pakax	jifkax	'opened eyes, will open eyes'
dakar		'stabbed'
zarak		'threw'

b. *Stops from historical geminates*

diber	'spoke'	< TH dibber
siper	'told'	< TH sipper
makar	'acquaintance'	< TH makkar
tabax	'cook'	< TH tabbax
tapil	'parasite'	< TH tappil

c. *Borrowings*

bilef	jebalef	'bluffed, will bluff'
pitpet	lepatpet	'talked a lot, to talk a lot'
mikroskop		'microscope'
zlob		'big, ungainly person'
fibrek		'fabricate'
falafal		'falafal'
fintez		'fantasised'
festival		'festival'

The /k/ in *pakax* (23a) is an important case in point, since it is a non-alternating obstruent co-occuring with an alternating one (p~f) in the same root. As Idsardi (1997:382) points out, this precludes an analysis in which certain roots are marked off as exceptional, and demands a phonological account in which some obstruents are alternating and some are non-alternating. Thus, in MH, we have a three-way distinction between /p/ specified in the lexicon as [-cont] (for *lepatpet* (23c)), /f/ specified as [+cont] (as in *fibrek* (23c)), and /P/ unspecified for [cont] in the lexicon (as in *pakax/jifkax* (23a)). Similarly we have /b/, /v/ and /B/ ([b~v]), and /k/, /x/ and /K/ ([k~x]). So we adopt the following constraints for MH:[5]

[5]Implicational constraints *Spirantisation* and *Stop* have been adopted here ahead of negative formulations (in prose "No stops following vowels"; "No fricatives") since implicational constraints succeed in filtering out the underspecified forms *Pakax* and *jiPkax* which would otherwise survive as the best outputs, incorrectly.

(24) IDENT([continuant])

Every value of [continuant] specified in the input is preserved in the output.

Spirantisation [-sonorant]→[+continuant] / V__

Obstruents are fricatives following vowels.

Stop [-sonorant]→[-continuant]

Obstruents are stops.

DEP([continuant])

Every [continuant] feature in the output must have an input correspondent.

Cumulative Order:

IDENT([continuant]), then *Spirantisation*, then *Stop*, then DEP([continuant])

Prespecified [-cont] or [+cont] is kept as it is by the IDENT constraint, while obstruents unspecified for [cont] are allotted [+cont] following a vowel by *Spirantisation*, or else [-cont] by *Stop*. These take effect one-by-one as in (25).

(25)

/jiPkax/	IDENT([cont])	*Spirantisation*	DEP([cont])	
jiPkax		*!		
jiPfax	*!			
jipkax		*!	*	
☞jifkax			*	

↓

/jifkax/	IDENT([cont])	*Spirantisation*	*Stop*	DEP([cont])
jiPkax		*!	**	
jipkax	*!	*	*	
☞jifkax			**	

↓

/jifkax/

/Pakax/	IDENT([cont])	*Spirantisation*	DEP([cont])	
☞Pakax		*		
Pafax	*!			
pakax		*	*!	
fakax		*	*!	

↓

/Pakax/	IDENT([cont])	*Spirantisation*	*Stop*	DEP([cont])
Pakax		*	*!*	
☞pakax		*		*
fakax		*	*!*	*

↓

/pakax/

When the *Spirantisation* constraint is added, [jifkax] can be selected over *jipkax*, and when the *Stop* constraint is added, [pakax] can be selected over *fakax*. Hence, any features not specified in the lexicon will take unmarked values, while IDENT([cont]) ensures that prespecified values, such as k in /Pakax/, will not change even if *Spirantisation* is violated as a result.

In addition to causing the retention of prespecified stops and fricatives, the pre-eminence of IDENT([cont]) also preserves spirantisation effects that are subsequently opaque. In forms like [bisfor] (21ii), *Spirantisation* first takes effect, but when the conditioning vowel is subsequently lost, the *Spirantisation* constraint has no force to retain the fricative in the absence of the vowel, and the more general *Stop* constraint favours *bispor* over *bisfor*. But *Stop* is prevented from taking effect, and *Spirantisation* overapplies.

(26) ***Why* [bisfor] *is not* *[bispor]**

bisaPor	Underlying Representation
↓	
bisafor	Spirantisation stage: /f/ in postvocalic environment
↓	
bisfor	Syncope stage: /f/ retained ahead of unmarked /p/ due to pre-eminence of IDENT([cont]) over *Stop*

This analysis also helps to explain the historical development of Hebrew. Tiberian and Modern Hebrew share this particular opacity effect whereby spirantisation overapplies with respect to syncope, fricative holding sway over stop. Hence, the IDENT([cont]) constraint is pre-eminent over the *Spirantisation* constraint in the phonology of Tiberian Hebrew also. Now, in Tiberian Hebrew, this leads to opaque specifications in the derivation but does not, as it happens, retain any lexical specification for there were none; all obstruents in the language alternated between stop and fricative. But because the continuant/noncontinuant distinction was made in the grammar by the pre-eminent IDENT([cont]) constraint, it would always be possible for a nonalternating stop or fricative to emerge in the language: if a specification for [continuant] were ever made in the lexicon, then it would be preserved specifically by the pre-eminent IDENT constraint. This prediction is borne out precisely by the borrowings and internal changes that have produced the expected nonalternating obstruents in Modern Hebrew.

The pre-eminence of IDENT([cont]) in a cumulative system secures stop realisations rather than fricatives in various other data cited by Idsardi (1997) and analysed in terms of counterfeeding. A fact about Spirantisation in Modern Hebrew is that it is restricted to labials and voiceless velars. That is, the spirantisation process does not produce *[θ ð ɣ], which are also generally absent from the Modern Hebrew consonant inventory. The obstruent inventory is given in (27), with the fricatives /f v x/ placed underneath their stop counterparts:

(27) *Modern Hebrew Obstruents*

p b t d k g s z ts ʃ

f v x

Markedness constraints *[coronal, +continuant, -strident] and *[dorsal, +continuant, +slack] affect both the inventory and the spirantisation process - a principled result in both source (markedness) and effect (wide-reaching restriction).[6] These constraints attain a pre-eminence in the grammar of Modern Hebrew they did not have in Tiberian Hebrew. The constraint against /ɣ/ groups Modern Hebrew with German, Russian and Polish with which Hebrew was in contact

[6] Idsardi (1997:377-9) attempts to derive the restriction on spirantisation from the inventory itself - in the rule-based framework by banning the rule from producing outputs not in the inventory list (the "Structure Preservation" condition), or in OT by creating highly specific and complex IDENT constraints such as one to hold identical the structure of /g/.

over the centuries, all of which have /f v x/ but not /ɣ/ (Idsardi 1997:377). The non-occurrence of /ɣ/ has an interesting effect in connection with Regressive Voicing Assimilation.

(28) Regressive Voicing Assimilation (Idsardi 1997:378)

a. [exzir] *[eɣzir] 'returned tr.'

[jixboʃ] *[jiɣboʃ] 'he will conquer'

b. /ja-gasos/ → [jiksos] *[jixsos] 'he will agonise'

/ja-gaxon/ → [jikxon] *[jixxon] 'he will lean'

In (28a), regressive voicing assimilation fails to voice the velar fricative in careful speech, respecting the constraint. In the examples in (28b), the /g/, which fails to spirantise according to the constraint, nevertheless devoices to [k] by regressive assimilation. However, although [k~x] alternations exist in Modern Hebrew, this [k] does not. In Idsardi (1997), the analysis depends on rule ordering, so that the spirantisation rule does not apply because it is ordered before the regressive voice assimilation rule in a counterfeeding relation. Once more, however, the prespecification analysis explains the facts. /g/ does not undergo spirantisation; it is non-alternating. Therefore, it will be prespecified in the lexicon:

According to Prince and Smolensky (1993) ... that underlying representation is chosen which leads to the violation of the fewest highly ranked constraints in the generation of surface form. Each morpheme has exactly one best underlying representation. As a result, there is no reason to worry about indeterminacy in underlying representation. ... Not surprisingly, FAITH constraints will always favor a fully specified form in the absence of higher-ranked constraints forcing underspecification. Thus for structures which exceptionally fail to show the expected alternations and instead maintain a constant surface form, Lexicon Optimisation will naturally cause them to be stored underlyingly in their surface form - that is, to be prespecified. (Inkelas, Orgun and Zoll 1997:410-411)[7]

---

[7]Idsardi (1997:382) questions Lexicon Optimisation, based on reports that mistakes and uncertainties occur among the three-way contrasts such as /k K x/ in MH. Thus, learners may make a non-alternating form alternating, or make an alternating form non-alternating. This does not falsify Lexicon Optimisation, however. Lexicon Optimisation is a linguistic principle which uniquely determines lexical representations from the surface representations. It does not say what happens when a human learner has been unable to establish a lexical representation from exposure to the language thus far, or is temporarily unable to access it, or otherwise chooses to construct another representation at the time of utterance. Such human factors inevitably superimpose on the basic linguistic system.

Either *ɣ or IDENT([cont]) would be sufficient to prevent /g/ from spirantising, but with regressive voicing assimilation (*RVA*), /g/ is devoiced to /k/, and it is IDENT([cont]) which continues to preserve the prespecified value [-continuant], so the /k/ fails to spirantise.[8]

(29) /ja-gasos/ ’he will agonise’

↓

/jigasos/	*ɣ	IDENT([cont])	*Spirantisation*
jiɣasos	*!	*	✓
☞ jigasos	✓	✓	*

↓

/jigasos/

↓

/jigsos/	*ɣ	IDENT([cont])	*Spir*	*RVA*
☞ jiksos	✓	✓	*	✓
jixsos	✓	*!	✓	✓
jiɣsos	*!	*	✓	*
jigsos	✓	✓	*	*!

↓

[jiksos]

---

[8]This provides a successful account of regressive voicing assimilation effects in careful speech. In fast speech, the *ɣ constraint is not respected, giving [eɣzir] ’returned tr.’ rather than [exzir], and [jiɣboʃ] ’he will conquer’ rather than [jixboʃ] (Idsardi 1997:378). This is attributable to the activity of regressive voicing assimilation in a stratum beyond that which defines the shape of words, dealing with utterance pronunciation. It is characteristic of processes of this kind that they are variably applied - hence the alternation between "careful speech" utterances and "fast speech" utterances. There is no uniform resolution one way or the other between (a) faithfulness to the citation form of a word ("carefulness"), and (b) reducing its phonological shape for convenience at speed.

Other counterfeeding effects may also be accounted for. A number of mergers have shaped the Modern Hebrew consonant inventory, /q/ > /k/, /ṭ/ > /t/, /ṣ/ > /ts/, /ħ/ > /x/, /w/ > /v/, /ʕ/ > /ʔ/, and geminates have been degeminated. Idsardi (1997) assumes that these are rules added to the end of the Tiberian Hebrew rules, so that spirantisation will not apply to the outputs of these rules because it is ordered earlier,[9] as in the following:

(30)	jiqra	dibber	
	------	-------	Spirantisation
	[jikra]	[diber]	Merger / Degemination
	'will read'	'spoke'	

The counterfeeding analysis suffers from the fact that the geminates and antiquated segments never surface, and the merger and degemination rules required for the analysis are not based on any observable processes in Modern Hebrew. They are rules of absolute neutralisation, merely devices to remove unattested phonemes and geminates. This is unnecessary abstractness, for if they are lexicalised in the form that is heard, they will be preserved in the surface form by the pre-eminent constraint IDENT([cont]) and will not be spirantised erroneously. In further data

---

[9] Idsardi (1997:381) states that the spirantisation rule is a "feature-filling-only" rule. This is necessary in order to account for stops in loanwords in Modern Hebrew, and in fact renders superfluous appealing to rule ordering to prevent spirantisation in other cases, since it operates in just the same way as the present analysis: spirantisation will not at any stage alter an obstruent already specified. Idsardi also applies the feature-filling condition to Tiberian Hebrew, in which case derivations such as rabb → rab → rav 'large (sg.)' (Idsardi 1997:368) are in need of revision. The reference to features with no specification that is necessary for defining feature-filling rules adds significantly to the formal complexity of rules (Inkelas, Orgun and Zoll 1997), and compared to the constraint-based analysis they are also an essentially stipulative move. The distribution of a feature may be predictable (i) in all instances in words (ii) in some instances and not others (ternary contrast) (iii) in some contexts and not others (neutralisation). The ternary contrast in MH requires feature-filling so as not to disturb the nonalternating obstruents, but a neutralisation pattern requires a feature-changing rule. But the situation in TH is case (i), predictability in all instances, and this is achievable with either type (though the default feature value must be by filling not changing). The historical facts of Hebrew merely force the conclusion that the feature-filling option would have to have been taken by stipulation. By contrast, a universal grammar employing violable IDENT(F) constraints provides precisely the desired distinction: a neutralisation pattern (iii) for feature F means IDENT(F) is violated in the language, but where there are patterns of

where spirantisation fails to apply, a productive alternation affects words such as those in (31): root-initial /n/ is deleted from the future forms on the right, and such /n/-less forms consistently fail to show spirantisation of the second root consonant, even though it does spirantise in the perfective forms on the left.

(31)	nava	'he derived from'	jiba	'he will derive from'
	nafal	'he fell'	jipol	'he will fall'
	nafax	'he breathed his last'	jipax	'he will breathe his last'
	nifkad	'he has been missing'	jipaked	'he will be missing'

Idsardi (1997) adopts a rule-based analysis with geminate formation followed by a free ride on the degemination rule /nP/ → /pp/ → /p/. The spirantisation rule does not apply because it is ordered before the degemination rule (though its feature-filling status also prevents it from applying). However, the degemination rule is a rule of absolute neutralisation not based on any observable process in Modern Hebrew. Again, an analysis in which IDENT([cont]) is pre-eminent extends to this data. Relevant here is that MAX([cont]) must stand pre-eminent over markedness constraints on [continuant] in the cumulative order in Hebrew, just as IDENT([cont]) does, for [continuant] features are retained in derivations, neither changing in value nor deleting. These two constraints act in concert here. Although /n/ is deleted, the [-continuant] feature from the underlying /n/ is preserved to the surface by MAX([cont]). It docks onto the second root consonant /P/ which, being an alternating obstruent, lacks a [continuant] feature of its own. Despite the fact that this consonant is postvocalic and therefore required to be a fricative by the

---

either total predictability (i) or ternary contrast (ii) for F, it is not violated. The naturalness of the TH-MH transition is confirmed since both sets of speakers would rank IDENT(cont) highly.

Spirantisation constraint, the pre-eminence of MAX([cont]) and IDENT([cont]) over *Spirantisation* secures a stop realisation at the expense of a *Spirantisation* violation.

In another pattern exemplified in (32), some plurals have postvocalic stops, even though the singular counterparts have fricatives:

(32)	kaf	'spoon'	kapot	'spoons'
	tof	'drum'	tupim	'drums'
	dov	'bear'	dubim	'bears'
	rav	'large sg.'	rabim	'large pl.'

Historically, the stops concerned were geminates, so in Tiberian Hebrew there was an alternation between intervocalic geminate stop and word-final fricative, e.g. *rav* / *rabbim*. Idsardi (1997:385) provides a complex rule-ordering analysis of (32) by positing underlying geminates for the Modern Hebrew forms. We could recast this in terms of a plausible constraint NoGeminates which, when added late in the cumulative order by the innovators of the change, suppressed the underlying geminates. However, the innovators may instead have simply failed to acquire geminates in their lexicon at all, and certainly modern learners exposed to data like (32) have no particular basis on which to set up geminates.

The alternation in (32) does not follow the usual stop / postvocalic fricative alternating pattern of Hebrew phonology; rather, what is displayed in (32) is irregular morphology. Thus, for a small class of words, plurality is marked by [-continuant] on the stem-final obstruent in addition to the suffix *-im*. Now the stops in alternations such as (32) are being lost from the language in favour of fricatives (Idsardi 1997:386). On our account, the fact and direction of this change are explicable. It is regularisation, a familiar process of morphological change, whereby the more widely-used pattern of plural-marking by

-*im* suffixation exclusively is being extended, and the uncommon added marking with stem-final [-continuant] abandoned.

Finally, Idsardi (1997) presents guttural deletion in Modern Hebrew and its effect on spirantisation. The gutturals /ʔ,h/ occur syllable-initially, but not syllable-finally, where they are deleted. For one class of glottal stops, deletion causes spirantisation of the following obstruent; in another class of glottal stops, spirantisation is blocked. Spirantisation is also blocked for deleted /h/.

(33)

			Spirantisation?	Origin
/ʔKl/[10]	jo**x**al	'will eat'	*Yes*	TH/ʔ/
/ʔBd/	ma**b**ad	'deed'	*No*	TH/ʕ/
/hPK/	ta**p**uxot	'instability'	*No*	TH/h/

The blocking of spirantisation here cannot be explained in the same way as in the case of deleted /n/ in (31), where the preservation of [-continuant] from the omitted /n/ phoneme ensured a stop rather than a fricative. Here we have deleted /h/ blocking spirantisation, which certainly supplies no [-continuant]. Instead, we adopt the constraint considered by Idsardi (1997) which rules out /ʔ,h/ in syllable coda by banning their superordinate feature [Guttural], placed after *Stop* in the cumulative order (34a). Then in the derivation of 'instability' (34b), deleted /h/ blocks spirantisation (obstruents are underlined at the stages where they are affected; obstruents that are already specified at an earlier stage are double-underlined):

[10]Despite the explicit proposal in Idsardi (1997) that alternating obstruents are underlyingly unspecified for [continuant], particular underlying forms were not given with the capitalised segments used in the text to signify the lack of full specification. We correct this here.

(34)

a. *NoGutturalCoda* *[Guttural] $]_\sigma$

Cumulative Order:

IDENT([cont]) before *Spirantisation* before *Stop* before *NoGutturalCoda*

b. /tahPuKot/

↓ *Spirantisation*

/tahPux̲ut̲/ /x/ spirantised, /t/ not spirantised since it is prespecified

↓ *Spirantisation, Stop*

/tahp̲ux̲ot̲/ /p/ stopped

↓ *Spirantisation, Stop, NoGutturalCoda*

/tap̲ux̲ot̲/ /h/ deleted; /p/ does not spirantise since already specified

‘instability’

We also require a way to distinguish the /ʔ/ of /ʔBd/, which blocks spirantisation, from the /ʔ/ of /ʔKl/ which does not. Idsardi (1997) distinguishes them in the lexicon by employing the original Tiberian Hebrew phoneme /ʕ/ along with /ʔ/ in underlying forms, which again we object to as abstract and unlearnable from exposure to Modern Hebrew. We propose that the lexical distinction in Modern Hebrew is between a glottal stop /ʔ/ and an underspecified guttural /?/, lacking [constricted/spread glottis] features. Where /?/ surfaces (in syllable onset positions), it is filled out by the grammar with the features of a glottal stop, the unmarked guttural. In coda positions, /?/ is deleted in the same way as /h/ was in (34), blocking spirantisation. /ʔ/ has the same distribution, but we delete it early from the coda to allow spirantisation.

An additional coda condition (35a) is required banning the glottal stop specifically, placed early in the cumulative order so as to filter /ʔ/ but not /?/. The derivations for ‘will eat’ and ‘deed’ in (35b) illustrate how spirantisation occurs in one case and not the other.

(35)

a. *NoʔCoda* $*ʔ]_{\sigma}$

Cumulative Order:

NoʔCoda before *Spirantisation* before *Stop* before *NoGutturalCoda*

b.	/jiʔaKal/	/ma?aBad/	
	↓	↓	*NoʔCoda*
	/joKal/	/ma?Bad/	/ʔ/ deleted
	↓	↓	*NoʔCoda, Spirantisation*
	/jox̲al/	/ma?Bad/	/x/ spirantised
	↓	↓	*NoʔCoda, Spirantisation, Stop*
	/jox̳al/	/ma?b̲ad/	/b/ stopped
	↓	↓	*NoʔCoda, Spir., Stop, NoGutturalCoda*
	/jox̳al/	/mab̳ad/	/?/ deleted, /b/ does not spirantise
	'will eat'	'deed'	

The loss of glottals from the coda by *NoʔCoda* is present in Tiberian Hebrew as well as Modern Hebrew, but general loss of Gutturals from the coda is an innovation. The cumulative order is in fact a reproduction of rule order in Idsardi (1997), Coda Glottal Deletion, Spirantisation, Stop

Formation, Coda Guttural Deletion, and the innovative element in both analyses comes at the end (of the rules, or of the active constraints).

This examination of the Modern Hebrew data raised in Idsardi (1997) demonstrates that underapplication can be described in CCT when the relevant Faithfulness constraint is pre-eminent over the relevant Markedness constraint. The presence of various postvocalic stops that fail to spirantise is analysed as prespecification either in the lexicon (exceptionality) or at a prior stage of the derivation (opacity), both of which are handled by Faithfulness over Markedness. More generally, Hebrew helps to illustrate the range of observable patterns that are available under Faithfulness over Markedness (F over M). In general, obstruents in the lexicon may or may not carry a [continuant] feature. If in some language all actual words with obstruents happen to carry a [continuant] feature, a simple stop:fricative contrast will be observed. Or if in all actual words with obstruents, none carries a [continuant] feature (as is claimed for Tiberian Hebrew), a predictable distribution of stops and fricatives will be observed. If some obstruents have a feature and some don't (as in Modern Hebrew), a three-way way contrast will be observed between stops, fricatives, and obstruents that alternate between the two. Such an alternation process shows underapplication. The situation with F over M contrasts with that of the opposite cumulative order, M over F, where outputs may sometimes be unfaithful, neutralising the relevant lexical distinctions. Distinguishing F over M from M over F gives rise to (36).

(36) **Hypothesis**

If a process underapplies (F over M) then it cannot neutralise (M over F);

if a process neutralises a contrast (M over F) then it cannot underapply (F over M).

We shall return to this hypothesis after incorporating some further underapplication phenomena into the theory.

6.2.3 More Underapplication: SOURCE constraints

We now consider evidence that derivations need to make continual reference to the underlying source, as well as to the form at a particular stage in the derivation. This comes from three further underapplication phenomena: *chain shifts*, *processes on derived structures*, and *subtractive morphology*.

In hypothesis (36) above, we have contrasted processes of neutralisation - which are predicted not to underapply, with processes of specification - which are predicted to underapply. Chain shifts are an additional phenomenon that do not fall within either of these types. Chain shifts have the form /A/→[B], /B/→[C], ... , where a series of underlying forms /A/, /B/, etc. are realised on the surface by the next in the chain. For example, vowels might raise to the next height level /a/→ε, /ε/→e, /e/→i. Chain shifts contain underapplication, since the process B → C does not apply to those instances of [B] that are raised from /A/. Now processes in chain shifts are not processes of specification - they alter underlying specifications. But they are not neutralising either - /A,B/→[C] is specifically avoided. Serial application of constraint hierarchies, as is used in Constraint Cumulation Theory, does not lend itself well to this. For if we apply constraint hierarchies cumulatively, at one step we will get /A/→/B/, but on the next step we may expect the next shift /B/→/C/ to be instantiated, erroneously giving overall /A/→/C/. Instead, the process B → C must be constrained to apply to underlying /B/ but not underlying /A/. This is the same problem as that pointed out by McCarthy (2000, 2002:159-162) for theories of *harmonic serialism*, in which outputs are re-submitted to a constraint hierarchy iteratively until no further change is possible (until "convergence").

The solution is that in addition to MAX constraints, that require maximum retention of features from one stage to the next in a derivation, we also need a family of constraints maximising preservation of features from the source form. We may call this constraint family

MAXSOURCE. We further assume the general ranking schema MAXSOURCE$_{\tau}$ >> MAX$_{\tau}$ for all constraint pairs refering to the same feature tier, or more generally to the same structural tier, τ. That is, preserving source structure is *more* important that preserving specifications at intermediate stages.

In (37), a chain shift is illustrated from a Spanish dialect. Vowels raise one level in height in the presence of the *-u* masculine singular suffix.

(37) Spanish dialect (McCarthy 2002:161, Hualde 1989)

	Stem	Fem Sg	Masc Sg	
a.	/nen/	néna	nínu	‘child’
	/sek/	séka	síku	‘dry’
b.	/gat/	gáta	gétu	‘cat’
	/blank/	blánka	blénku	‘white’

In (37a), mid vowels raise to high, in (37b), low vowels raise to mid. The chain shift generalisation is that on some particular scale (here, vowel height), there may be one MAXSOURCE violation but there cannot be two for the same segment in the same derivation. This is expressed by the additional device of local conjunction, as shown by Kircher (1996). Thus MAXSOURCE([high]) may be violated (37a), MAXSOURCE[low]) may be violated (37b), but the conjoined constraint [MAXSOURCE([low])-&- MAXSOURCE([high])]$_{Seg}$ which measures violations of the two conditions within a single segment, remains intact, since only one feature, never both, is removed from the vowel.

In (38), we demonstrate how the features of *-u* spread. We use the abbreviation MS for MAXSOURCE; and ALIGNHEIGHT = ALIGNLEFT([+high],Word), ALIGNLEFT([-low],Word).

(38)

		MS([high])-&-MS([low])	ALIGNHEIGHT	MS([high])	MS([low])
/nen+u/					
	nenu		*!		
	☞ninu				*
/gat+u/					
	gatu		**!		
	☞getu		*		*
	gitu	*!		*	*

(39) shows what happens at any subsequent stage. The value of MAXSOURCE constraints is that, in the subsequent accumulation of constraints at further steps, **gitu* can never be derived from /getu/ from underlying /gatu/ because the conjoined constraint still disallows changes in both [low] and [high] to the original /a/.

(39)

/getu/		MS([high])-&-MS([low])	ALIGNHEIGHT	MS([high])	MS([low])
	gatu		**!		
	☞getu		*		*
	gitu	*!		*	*

↓

/getu/

The wider effects of source constraints include processes which affect only *derived* feature specifications without affecting underlyingly specified forms. This is given by the cumulative scheme MAXSOURCE *before* Markedness *before* MAX. This occurs in the case of depalatalisation before coronals in Slovak: “a derived /l´/ undergoes depalatalisation but an underlying //l´// does not” (Rubach 1993:268). The paradigms in (40) show contrast between: (a) conditioned palatalisation (before front vowels) which is depalatalised before a coronal (but not a labial) in the genitive singular nominal; and (b) underlying palatalisation which is not depalatalised:

(40) Depalatalisation in Slovak (Rubach 1993)

a. /strel/	[strela] ‘shot’		
	[strel´ets] ‘shooter’		
	[streltsa] ‘shooter’(gen.sg.)		[strel´ba] ‘shooting’
b. /topol´/	[topol´] ‘poplar’		
	[topol´ets] (dimin.)		
	[topol´tsa] (dimin. gen sg.)		

In addition, some words like *bol´ʃevik* ‘Bolshevik’ have an underlying //l´// which is preserved. As well as processes which affect only derived structure, there are other phonological processes which take effect only in *derived environments* where the process only applies when the structure that triggers it, as opposed to the structure it affects, is specified in the course of the derivation and not in the underlying root. This is an area open for further investigation, but the SOURCE-constraint approach may be expected to absorb current proposals for describing derived-environment processes using Faithfulness constraints (Łubowicz 1999), just as proposals for accounting for chain shifts were absorbed above.

Subtractive morphology is a phenomenon where the underlying source is again significant. In subtractive morphology, a grammatical meaning is marked not by an affix but by an alteration to the stem. In that case, a MAXSOURCE constraint is violated – minimally – due to another constraint requiring the marking of some morphological category by a special deviation from the underlying source. Consider the case of Lardil nominals (Hale 1973, Itô 1986, Wilkinson 1988, Prince and Smolensky 1993) exemplified in (41). In the nominative, there is no affix, and the nominative case is characterised instead by the removal of final vowels (41a,b,c).

(41) Lardil nominals

	Stem	Nominative	Nonfuture Accusative	Gloss
a.	/yiliyili/	yiliyil	yiliyili-n	oyster sp.
b.	/mayařa/	mayař	mayařa-n	rainbow
c.	/yalulu/	yalul	yalulu-n	flame
d.	/kentapal/	kentapal	kentapal-in	dugong
e.	/ŋaluk/	ŋalu	ŋaluk-in	story
f.	/muŋkumuŋku/	muŋkumu	muŋkumuŋku-n	wooden axe
g.	/pulumunitami/	pulumunita	pulumunitami-n	young fem dugong

Nominative marking does not extend to deletion of consonants (41d), although final consonants are vulnerable to deletion on phonological grounds (41e), if they do not satisfy the restrictions on syllable coda. In Lardil, codas admit only nasals homorganic to a following onset or apical sonorants. This phonological deletion of consonants, which may accompany the deletion of the final stem vowel, tends to expose new vowels to final position in nominative forms (41f,g). The freshly exposed vowels are not deleted, only the truly stem-final ones, an apparent underapplication of final vowel deletion.

In the analysis of Prince and Smolensky (1993), a brute force constraint is placed in the hierarchy of constraints to enforce the nominative marking on the final vowel. Here, we adopt the generalised REALISEMORPHEME constraint (Kurisi 2001).[11] In direct contrast to MAXSOURCE constraints, this constraint requires a deviation (any deviation) from the underlying source, specifically in order that a morphological category be marked. In (42), this constraint is ordered prior to the relevant MAXSOURCE constraint in order to force a realisation of the nominative case at the expense of the retention of the final vowel.

(42) REALISEMORPHEME

Every morphological feature present has a reflex in the output.

Cumulative Order: REALISEMORPHEME *before* MAXSOURCE-V

/muŋkumuŋku/

	CODA COND	REALISE MORPHEME	MAX SOURCE-V	MAX-V
.muŋ.ku.muŋ.ku.		*!		
.muŋ.ku.muŋk.	*!		*	*
.muŋ.ku.muŋ.	*!		*	*
☞.muŋ.ku.mu.			*	*
.muŋ.ku.			**!	**

↓

/.muŋ.ku.mu./

At subsequent steps in the cumulation of constraints, no further vowels are deleted or inserted, as shown in (43). Deletion of a further vowel is held back both by MAX-V and MAXSOURCE-V. An

[11] This approach contrasts with an alternative in which the subtracted form is measured against other members of the paradigm (Alderete 2001).

inserted vowel that restores identity with the underlier under a natural correspondence relation,[12] so satisfying MAXSOURCE-V, would nevertheless violate both DEP-V and REALISEMORPHEME.[13]

(43)

/muŋkumuŋku/

↓

...

/.muŋ.ku.mu./

	DEP-V	REALISE MORPHEME	MAX SOURCE-V	MAX-V
☞.muŋ.ku.mu.			*	
.muŋ.ku.			**!	*
.muŋ.ku.muŋ.ku.	*!	*!		

↓

/.muŋ.ku.mu./

In a theory of violable constraints that requires violation to be minimal, the conflict between REALISEMORPHEME and MAXSOURCE constraints over identity or non-identity with the underlying source is resolved by at most one alteration to the stem, not two or more. It has been

[12]Extending use of the Correspondence assignment function, given for input-output pairs in chapter 2, I assume that, at each stage, potential output structures are assigned correspondence relations with (i) the form from the previous stage (ii) the underlying form. As shown in chapter 3, it is a corollary of the minimal constraint violation metric that a *natural* correspondence mapping is the most harmonic, hence natural correspondences are the ones considered.

[13]The basis for the pre-eminence of DEP-V over REALISEMORPHEME lies in other Lardil nominals such as *wiʈe* **wiʈa*, where the expected *a*-epenthesis seen elsewhere in /yak/→*yaka* is not used. In this way, the stem is held identical at the expense of nominative case marking on such forms (Prince and Smolensky 1993, chapter 7).

claimed (Kurisi 2001) that this restriction on morphological marking on the stem holds across languages generally.

In this subsection, we have described further underapplication phenomena using constraints that require preservation of the underlying source: chain shifts, processes on derived structure, and subtractive morphology. Note that MAXSOURCE constraints serve a different function to MAX and DEP constraints, despite their overlapping effect: MAXSOURCE maximises the preservation of source material, while MAX and DEP have an operational function to minimise the differences between one stage of the derivation and the next (whether constructive or destructive). There is no DEPSOURCE constraint family to complement MAXSOURCE, since it would serve neither function. Precisely the same functional distinction is recognised for phonology by Paradis (1988, 1996, 1997) in her Preservation Principle and Minimality Principle. This position has further consequences in CCT, which we now examine.

6.2.4 A Hypothesis

Source constraints aim to preserve underlying specifications, so where some markedness constraint is pre-eminent over MAXSOURCE, it forces neutralisation of underlying specifications, but where MAXSOURCE is pre-eminent over a markedness constraint, the markedness constraint cannot force processes in contravention of underlying specifications, and underapplies in respect of them. This dichotomy between neutralisation and underapplication only repeats hypothesis (35) made in connection with Modern Hebrew. But there is a further consequence now. Neutralisation processes – or rather, processes caused by markedness constraints dominating all relevant faithfulness constraints – cannot *over*apply any more than they underapply.

To see this, consider a hypothetical example. Suppose all word-final obstruents devoice in a language which disallows labials in syllable coda, saving final labials by epenthesis to place them in the onset. Then for an underlying form like */kob/* i.e. with a final voiced labial, consider the two possible outputs */kobe/* and */kope/*. We certainly expect epenthesis, but the question is whether devoicing will occur. Suppose the constraint on voiced final obstruent is pre-eminent over the coda condition. Then we have the first step */kob/→/kop/*. When the coda condition is added, leading to epenthesis, */.ko.pe./* and */.ko.be./* will both satisfy the coda condition, and *both* will satisfy the proscription against final voiced obstruents. Markedness does not decide between them, so it is down to faithfulness. /.ko.pe./ violates MAXSOURCE for voicing, since the source is /kob/, but /.ko.be./ is faithful to the source. There is a general prediction here: once a pre-eminent markedness constraint no longer discriminates between the two alternants because the context has changed, the relevant MAXSOURCE constraint favours a reversion to the underlying value. MAX and DEP constraints favour retaining the overapplied value from the previous derivational stage, but if, as we have assumed, MAXSOURCE$_\tau$ is pre-eminent over MAX$_\tau$ and DEP$_\tau$ for each structural tier τ, restoration of the underlying value prevails and overapplication is rendered impossible. To demonstrate the generality of the result, an algebraic formulation is given in the box at the end of this section.

The prediction has currency in the empirical record. Neutralisation processes consistently fail to overapply or underapply. In Lithuanian (Kenstowicz and Kisseberth 1973:6), regressive voicing assimilation applies in consonant clusters, e.g. [ap]-*arti* 'finish ploughing' vs. [ab]-*gyventi* 'inhabit', but not to homorganic consonant clusters, which are broken up by epenthesis, e.g. [ap]-*i-begti* 'to run around'. Thus the process refrains from overapplication. Indeed, voicing assimilation never applies to consonants separated by epenthesis in any known language where the two processes co-exist (Kenstowicz and Kisseberth 1977:173). Several more examples of neutralisations and deletions which fail to overapply are quoted by Kenstowicz and Kisseberth (1973): Tübatulabal stop devoicing; Klamath preconsonantal neutralisation in glottalisation, aspiration, and voicing; Takelma preconsonantal devoicing/deglottalisation; Lithuanian metathesis and degemination; Yawelmani Yokuts apocope and vowel shortening; Washo vowel drop and stress reduction; West Greenlandic Eskimo metathesis and high vowel lowering. Similarly, in Southeastern Tepehuan (Kager 1997), three neutralisation processes - intervocalic mutation of /v/ to [p] in reduplicated stems, conversion of voiced obstruents to pre-glottalised nasals in syllable coda, and complete assimilation of coda /h/ to the following consonant - are transparent, failing to overapply or underapply where they might conceivably have done when the surrounding vowels are deleted.

In contrast, observe that familiar examples of overapplication in phonology are those which instead *insert* material, and which thus offend no MAXSOURCE constraints. One example is Hebrew spirantisation which inserts the [+continuant] feature. The conditioning environment is a preceding vowel, but the process survives deletion of some of the vowels, an overapplication (see 6.2.2 above). In other cases, epenthesis breaks up consonant clusters of which one consonant may be subsequently deleted: examples include Eastern Massachusetts English fiːr → fiːjə, also Hebrew deʃʔ → deʃe (McCarthy 1999a) and Turkish inek-m → ine-i-m (Orgun and Sprouse 1999). Broadly, then, we have the following distinction:

(44)

Transparent Application	Likely To Overapply or Underapply
Neutralisation	Conditioned Variation
Deletion	Insertion

This dichotomy recalls the debates over rule ordering in phonology in the 1970s. The maximum utilization principle (Kiparksy 1968) which favours overapplication, and the minimisation of opacity principle (Kiparsky 1976) which favours transparency, fell foul of different examples. Kenstowicz and Kisseberth (1973, 1977) noticed that there were interesting subregularities, but articulated difficulties that held everyone back from finding a solution in the rule-based approach of the day:

> The generalization that seems to emerge from the examples discussed above is that rules of assimilation, neutralisation, etc., tend to be predictably ordered to apply to "surface" rather than "abstract" syllable structure. .... However, when one begins to examine this "generalization" a number of difficulties immediately spring to mind. To cite just one example, many languages have the following two rules in their grammars: a rule palatalizing a consonant before i, and a rule of apocope. Given the claim that bleeding orders are expected between an assimilation rule and a rule affecting syllable structure, we would predict that the apocope rule should precede, and, hence, bleed the palatalization rule. Yet a cursory inspection of languages possessing these two rules indicates that a non-bleeding order is typical. Perhaps one might suggest that a non-bleeding order is preferred because the i which drops by apocope leaves a "trace" on the preceding consonant. We have not as yet examined the matter sufficiently to be able to determine if this suggestion is at all viable. (Kenstowicz and Kisseberth 1973:10)

Taking up these concerns, it is clear that assimilation to segments which are subsequently deleted may or may not go against the pattern in (44) - assimilation can be a form of neutralisation or of conditioned variation. Palatalisation may involve conditioned variation between assimilated and default allophones [t~c] or [k~k̟]. However, if palatalisation, for example, brings about a neutralisation of contrast before /i/ between alveolar and palatal /t/:/c/ or velar and palatal /k/:/c/, and that /i/ is lost to apocope, then we have something which goes against the pattern in (44). Nevertheless, these *stability effects* are readily accommodated, given that contemporary phonological theory provides us with autonomous, spreadable features and Faithfulness constraints that can require the retention of features even when the segment they come from is deleted. The stability of the lateral gesture in the morphophonemic alternation of Klamath nl̥→lh due to MAX([+lateral]) illustrates in (45). This was first quoted in 1.2.1. In CCT, at the point where the constraint OCP {sonorant} is added, MAX([+lateral]) ensures that /lh/, which retains the lateral feature, is selected over /nh/.

(45)

/nl̥/

	OCP {sonorant}	MAX ([+lateral])	MAX ([+nasal])	DEP (Assoc)	MAX (Assoc)
nl̥	*!				
ll̥	*!		*	*	
☞lh			*	*	*
nh		*!			*

↓

/lh/

In CCT (unlike standard Optimality Theory) this strategy is available within a system that also allows for overapplication of insertive processes in serial interaction with other processes, thereby predicting opacity in all the places where it is found, and solving Kenstowicz and Kisseberth’s puzzle.

Stability effects are of a kind with both *prespecification* and *chain shifts*, since in each case a pre-eminent Faithfulness constraint prevents a fully transparent solution. We may sum up this result as in (46):

(46) **Transparency of Pre-Eminent Markedness**

Markedness constraints that are pre-eminent over Faithfulness constraints for the affected material will be fully transparent in their effect, triggering structural alternations *when and only when* failure to do so would result in surface violation.

Non-Transparent Cases:

i. **Prespecification**: pre-eminent Faithfulness constraint forces retention of feature in direct conflict to Markedness.

ii. **Chain Shifts**: pre-eminent Faithfulness constraint excludes multiple changes to single phoneme in conflict to Markedness.

iii. **Stability Phenomena**: a pre-eminent Faithfulness constraint salvages a feature from a deleted segment, causing a mutation to the adjacent segment whose cause is of course absent.

Except for stability effects, neutralisations and deletions are predicted to be transparent, as in (47):

(47) **Transparency of Neutralisation Hypothesis**

Contextual neutralisation and elision of phonemes occurs if and only if the context in which it occurs is present in the actual surface representation, except when neutralisation is caused by assimilation to a phoneme that is deleted.

(47) is a prediction of Constraint Cumulation Theory on the descriptively necessary assumption that MAXSOURCE constraints are present, and on the further assumption that they are pre-eminent over their MAX and DEP counterparts. Constraint Cumulation Theory is a theory of Markedness-Faithfulness interaction which accurately predicts which processes must be fully transparent and which may be opaque.

Nullification of Overapplication by MAXSOURCE

Consider structures p_0, p_1, p_2, p_{12}, where p_0 is (isomorphic to) the underlying source form, and there is a disparity between p_1 and p_0 violating MAXSOURCE1 and MAX1, a disparity between p_2 and p_0 violating MAXSOURCE2 and MAX2, and both disparities between p_{12} and p_0. Let there be a Markedness constraint M_a which triggers $p_0 \rightarrow p_1$. That is, p_0 violates M_a , and M_a >> MAXSOURCE1 >> MAX1. Let there be a Markedness constraint M_b which triggers $p_0 \rightarrow p_2$ & $p_1 \rightarrow p_{12}$. That is p_0, p_1 violate M_b, and M_b >> MAXSOURCE2 >> MAX1.

Let M_a >> M_b. With M_a and M_b so defined, and without MAXSOURCE, this would be sufficient to trigger an overapplication of M_a, $p_0 \rightarrow p_1 \rightarrow p_{12}$.

At the first step: M_a triggers $p_0 \rightarrow p_1$ as normal

On the second step with M_a & M_b, M_b disfavours p_0 and p_1;

p_2, p_{12} both satisfy M_b and violate MAXSOURCE2;

p_{12} additionally violates MAXSOURCE1;

p_{12} is less harmonic (unless there is some saving constraint dominating MAXSOURCE1.

Hence, p_2 wins.

6.2.5 Transparency of Neutralisation: Yokuts

A more complex pattern, that of vowel qualities in Yawelmani Yokuts, provides a further significant test of the Transparency of Neutralisation Hypothesis. We begin straightforwardly. High vowels are given up in favour of mid when long, as in (48).

(48)	a.	meek’-i-t	'was swallowed'	cf. mik’aaʔan 'is swallowing'
	b.	ʔooṭ’-u-t	'was stolen'	cf. ʔuṭ’aaʔaanit 'is to be stolen'

Next, the application of mid height overapplies. The syllable inventory of Yokuts lacks closed syllables with long vowels, but when roots such as those in (48) come in closed syllables, the mid vowels are retained under shortening, as shown in (49).

(49) a. mek-k’a 'swallow!'

b. ʔoṭ’-k’a 'steal!'

Previous accounts (Kuroda 1967, Kisseberth 1970, Kenstowicz and Kissebeth 1979, Archangeli 1983, 1985, 1988, 1991, Steriade 1986, Mester 1986, Archangeli and Suzuki 1997, McCarthy 1999a) have always held that there is neutralisation in height on long vowels, so-called "High Vowel Lowering". In Constraint Cumulation Theory, however, assuming the presence of MAXSOURCE constraints, the height alternation cannot be a neutralisation for then by Transparency of Neutralisation it would be incapable of overapplying. We would have a derivation like (50).

(50)

/miikk’a/

	NOLONG HIGHV	MAXSOURCE ([high])	MAX ([high])
.miik.k’a.	*!	✓	✓
☞.meek.k’a.	✓	*	*

↓

/.meek.k’a./

	NoLong HighV	Foot-Binarity	MaxSource ([high])	Max ([high])
.miik.k’a.	*!	*	*	*
.meek.k’a.	✓	*!	✓	✓
.mek.k’a.	✓	✓	*!!!!	✓
💣.mik.k’a.	✓	✓	✓	*

↓

*/.mik.k'a./

The first step *.Ci:C.* → *.Ce:C.* is predictable provided that the proscription against long high vowels is pre-eminent over MaxSource([high]). However, on the second step when the vowel is shortened, *.CiC.* eclipses *.CeC.* because it reproduces the underlying height: although Max([high]) favours retention of the lowered value of the previous stage, MaxSource([high]) prefers the output with a high vowel matching the underlying form. The markedness constraint NoLongHighV that prompts lowering at the previous stage is no longer pressing for a mid vowel, being vacuously satisfied because the vowel is short.

We conclude that since *.CeC.* shows up in the Yokuts data, it must be that there is no such underyling [+high]: instead there is underspecification. Then specification of [-high] on long vowels can overapply along the lines of (51):

(51)

/mIIkk’a/

	NoLong HighV	MaxSource ([high])	Max ([high])
.miik.k’a.	*!	✓	✓
☞.meek.k’a.	✓	✓	✓

↓

/.meek.k’a./

	NoLong HighV	Foot-Binarity	MaxSource ([high])	Max ([high])
.miik.k’a.	*!	*	✓	*
.meek.k’a.	✓	*!	✓	✓
☞.mek.k’a.	✓	✓	✓	✓
.mik.k’a.	✓	✓	✓	*!

↓

/.mek.k’a./

Re-analysing the pattern as conditioned variation of height, to be handled by the insertion of [-high] and [+high], is analogous to the conditioned variation in Hebrew of stops and fricatives, handled by the insertion of [+continuant] and [-continuant]. Like the Hebrew pattern, the Yokuts vowel height alternation exhibits opacity. However, there is a further complexity in the Yawelmani Yokuts vowel pattern.

Vowel height is a crucial factor in vowel harmony in Yawelmani. If vowels agree in height, then they become completely identical. Three particular suffixes serve for expository examples in (52), but the effect extends to all suffixes. If the root vowel (if there are two they will be identical) and suffix vowels are both high, then suffix vowels will take on all the features of the stem vowel, as in (52a,c). If stem vowel and suffix vowels are both non-high, then suffix vowels will take on all the features of stem vowels, as in (52b).

(52) a. -hIn, aorist I = high, realised {[i] or [u]}

xil-hin	’tangled’
hud-hun	’recognised’
c’om-hun	’destroyed’
xat-hin	’ate’

b. -Al, dubitative A = non-high, realised {[a] or [o]}

xat-al	’might eat’
bok’-ol	’might find’
k’oʔ-ol	’might throw’
hud-al	’might recognise’

c. -sIt-, indirective

t’ul-sut-hun	’burns for’
bok’-sit-k’a	’find (it) for!’

Although vowel harmony is pervasive in the language, its dependence on vowel height means that it is interrupted by the alternations in vowel height mentioned earlier:

(53)	a.	meek’-i-t	’was swallowed’	cf. mik’aaʔan ’is swallowing’
	b.	ʔooṭ'-u-t	’was stolen’	cf. ʔuṭ'aaʔaanit ’is to be stolen’
	c.	c’oom-al	’might destroy’	
	d.	doos-i-t	’was reported’	

In (53a), the mid quality of long vowels is shown in a stem vowel which otherwise appears high (i.e. when it is not long). Alternating mid vowels behave as if high as far as vowel harmony is concerned: in (53b) spreading roundness to a high affix vowel, but in (53c), an alternating mid vowel does not spread to a non-high affix vowel. Note the contrast in behaviour between /o/ in

(53b), which behaves as high with respect to harmony, with /o/ in (53d), which does not. /o/-vowels that are not height-alternating as in (53d) are exceptional and will be prespecified with [-high]. Vowel Harmony, then, displays underapplication and overapplication with respect to Lowering: non-high affix vowels will not harmonise with lowered root vowels, and high affix vowels retain their harmony with lowered root vowels. Our established response is to say that roundness is conditioned: conditioned feature values are not specified underlyingly but filled in by the grammar. Then overapplication is expected, because it cannot be nullified by MAXSOURCE if there are no underlying specifications.

The difficulty is that roundness is conditioned by height, but height itself is also conditioned, and therefore not present underlyingly. If there is no underlying [+high], then it is difficult to see what the second vowel in *ʔooṭ'-u-t* can harmonise with. The specifications left undetermined by the grammar are now in (54).

(54)

Base Vowels	[low]	[high]	[round]	Affix Vowels	[low]	[high]
I [i~u]	-		-	V [i~u~e~o]	-	
U [u~o]	-		+	A [a~o]		-
o	-	-	+			
a	+	-	-			

Cases like *ʔooṭ'-u-t* have been taken in the literature to show that such vowels are underlyingly [+high] and change to [-high] once vowel harmony has been instantiated (ʔuuṭ'-i-t → ʔuuṭ'-u-t → ʔooṭ'-u-t). The alternative analysis is that [+high], rather than being prespecified underlyingly, is temporarily inserted in the derivation, as a default. The default nature of [+high] has crosslinguistic support (Archangeli 1988). Its insertion will be required by the well-formedness constraint "every vowel must be associated to a [high] feature" (WF). WF recalls other well-formedness requirements in autosegmental phonology, e.g. "every tone-bearing unit must be associated to a tone" (Goldsmith 1976) which we assume generalises to any two tiers whose elements associate. [+high] is made the default by utilising the claim that phonological constraints refer to marked structure (Inkelas 1994, following a proposal given by Paul Kiparsky at the TREND conference in 1994). Then we have Faithfulness constraints DEP([high]) and the more specific DEP([-high]), MAX([high]) and the more specific MAX([-high]), which make reference to the marked [-high]. This predicts, correctly, that unmarked features are inserted more generally and deleted more generally than marked ones cross-linguistically.[14]

This leads to the following derivations (I,U,V are as defined in (54): I=[-low,-round], U=[-low,+round], V=[-low]):

(55) Constraints in Cumulative Order:

MAXSOURCE([high]) *before* WF *before* Harmony *before* NoLongHighV *before* FootBinarity *before* MAX([high]) *before* MAX([-high]) *before* DEP([high]) *before* DEP([-high])

[14]This proposal replaces that of context-free markedness constraints of the form *[m feature], e.g. *[-high]. These have persisted in the OT literature (Kager 1999, McCarthy 2002) despite the correct prediction of F([m feature]) constraints cited in the text. Here in Yokuts, a context-free markedness constraint *[-high] would fail, since in order to insert default high vowels it would have to dominate the later-added NoLongHighV, and would therefore wrongly suppress the formation of long mid vowels.

a.	ʔUUț'-V-t	c'UUm-hVn	
	↓	↓	*WF*
b.	ʔuuț'-I-t	c'uum-hIn	All vowels given [+high] by default
	↓	↓	*WF; Harmony*
c.	ʔuuț'-u-t	c'uum-hun	Suffix vowels harmonise with root vowels
	↓	↓	*WF; Harmony; NoLongHighV*
d.	ʔooț'-u-t	c'oom-hun	Root vowels lower; suffix vowels retain [+high]
	↓	↓	*WF; Harmony; NoLongHighV; FootBinarity*
e.	ʔooț'-u-t	c'om-hun	*c'om* vowel shortened, but retains mid-height

As the Well-Formedness constraint is added, the least costly way of satisfying it is to add [+high] features, avoiding violation of DEP([-high]) (55b). Because the vowels share the same height, Harmony attributes [+round] to them both too (55c).[15] Harmony now overapplies with respect to Lowering at (55d) because when the root vowels lower to satisfy NoLongHighV, [+high] is retained by the suffix vowels due to the influence of MAX([high]). Finally, Lowering overapplies again with respect to shortening, with MAX([high]) retaining the /o/ in /c'omhun/.[16]

In Constraint Cumulation Theory, the marked Faithfulness constraints influence every stage of the derivation, since Faithfulness constraints are present at every stage, so when a well-formedness constraint calls for a feature, it is the unmarked one that will be provided. An

[15]According to the representational system of Mester (1986), harmony follows from the constraint OCP([high]). In fact, the order of WF and OCP([high]) is not crucial: they might both be added at the same step and determine vowel features in parallel.

[16]This is an unnatural derivation in the sense defined in chapter 3. The [+high] is not in the underlying form or the surface form, and therefore invisible in the underlying-surface correspondence mapping. It is only detected indirectly by its influence on other vowels in vowel harmony. It has been claimed by Steriade (1986) and Mester (1986) that Harmony and Lowering belong to different strata of the grammar, one applying after the other. These 'strata' correspond to subportions of the derivation which are themselves natural (∅→[+high] and [+high]→[-high]), and each portion produces well-formed vowel-height association, whether [+high] or [-high].

analysis with a temporary insertion of the default [+high] during the portion of the derivation where vowel harmony is determined formally reflects the fact that long mid vowels behave as high for purposes of harmony. In fact, the presence of marked Faithfulness constraints predicts that only unmarked features can be inserted temporarily in a derivation and have opaque effects such as that seen in Yokuts: this is an *opaque emergence-of-the-unmarked* effect. In this way, the added effect of vowel height on vowel harmony, which involves overapplication and underapplication, still submits to an analysis in which the [high] feature is specified by the grammar, as the Transparency of Neutralisation Hypothesis (47) predicts.[17]

The analysis brings Yawelmani Yokuts into line with the underspecification / prespecification approach developed in Inkelas (1994) and Inkelas, Orgun and Zoll (1997), taken up here in Constraint Cumulation Theory to derive new predictions about opaque interactions. Following Inkelas (1994), the underspecification employed is based not on principles governing lexicon forms, but on the actual surface alternations. The underlying forms are ones that are consistent with the Markedness-Faithfulness interactions that describe the alternations, and are not based on direct constraints on the input or on the lexicon, in accordance with the Richness Of The Base principle, which is Optimality Theory's legacy in our understanding of the lexicon.[18] It is also a feature of our analysis that prespecification is possible, creating exceptions to the effect. This prediction too is borne out by instances of /o/ such as /doosit/ which, being specified as

---

[17] Another problematic aspect of the vowel height pattern in Yawelmani Yokuts noted by Sprouse, Inkelas and Orgun (2001) produces another unnatural derivation, which appears to go against the tight Duke of York limitations in CCT given in 6.1.2, and it is difficult to see whether it may submit to a temporary-default treatment. *sognut* is derived from underlying /sugnit/, where lowering of the first vowel crucially depends on morphologically-conditioned vowel length at an intermediate stage before shortening to fit syllable structure, /u/ → u: → o: → o .

[18]Archangeli and Suzuki (1997) propose that the *input markedness* constraint, $V \approx [+high]_I$ "Input vowels must be [+high]" is necessary to ensure that Yokuts [e:] is lexicalised /i/ even in cases where it never surfaces high, so that it will be predicted to behave as a high vowel for purposes of vowel harmony. In doing so they abandon the Richness of the Base principle, claiming instead that constraints on the input mirror the output markedness constraints. This must be combined with two further technical innovations, *disparate correspondence* constraints, and the *input-else* constraint condition, to complete their account of opacity in the Yokuts vowel system. This is an unwelcome proliferation of devices, and, as has been shown by Idsardi (1997) for Hebrew, the approach fails to provide an adequate account of the historical development of patterns it describes.

[-high], does *not* behave as high with respect to harmony. However, McCarthy (1999a) emphasises the lack of an /e/ in Yokuts to complement the /o/. He therefore proposes an alternative analysis in which vowels in Yokuts participate in a chain shift that steps down in height i:→e(:)→a(:), and uses this to explain the lack of /e/, unlike /o/, in underlying forms: [e]/[e:] will be lexicalised as /i/, since /e/ is realised as [a(:)] by the chain shift. Nevertheless, the naturalness of the chain shift analysis is compromised by several caveats: *short /i/* is specifically excluded; back vowels also lower, but only from high to mid and *not further*; and in the Wikchamni dialect of Yokuts, a *raising* process takes short mid vowels in the opposite direction. This complexity is avoided on our account by the use of the single constraint NoLongHighV, but with pre-eminent Faithfulness, creating conditioned variants [i]/[e:] [u]/[o:] which admit – as expected – both exceptions and opacity. We have not explained the lack of non-alternating /e/ (or /i/ or /u/) on our account, but it may well be right *not* to attribute the lack of /e/ to the grammar, since the explanation may lie in language learning and use, as follows. Phonetic [o], when followed by the frequent vowel [i] as in [doosit], *must* be /o/ and not the underspecified vowel /U/ underlyingly because if it was /U/, it would have caused rounding to the [i]. This reinforces the lexicalisation /o/. By contrast, [e] when followed by the same frequent vowel [i] looks as if it *could* be a lowered form of /I/, as it is in [meek'it]. The fact that the vowel harmony pattern reinforces /o/, but not /e/, suggests that /o/ will be more stable.

6.3 Conclusion

Constraint Cumulation Theory is a synthesis of serial and hierarchical interaction between phonological constraints. Its predictions match the empirical record at many points, deriving patterns of overapplication, mutual interdependence, default, reprosodification, reversal in absence of conflict, prespecification, chain shifts, processes confined to derived structures, subtractive morphology, stability effects, and multiple overapplication.

These depend on an interaction between Markedness constraints, added cumulatively, and Faithfulness constraints which not only regulate each step of the derivation but also measure the retention of underlying specifications. Along with this, the twin strategy of underlying specification versus default specification (Zwicky 1986) is harnessed: contrastive features are specified underlyingly, while features whose values alternate are not, allowing default values to be inserted. This approach has been tested against the more complex patterns in Modern Hebrew and Yawelmani Yokuts, in addition to a range of other examples.

The cumulation of constraints at successive constraint evaluations contrasts with an approach in which iterative constraint evaluation uses the same constraint hierarchy repeatedly ("Harmonic Serialism", Prince and Smolensky 1993), and CCT's wide descriptive coverage dwarfs the limited capacity of this system (see McCarthy 2000). Another alternative, Derivational Optimality Theory (Rubach 1997, 2000), uses slight variations in the constraint hierarchy over two or three iterations in order to account for complexities in Slavic languages, but lacks the general prediction of Transparency of Neutralisation that comes from the cumulative interaction of constraints constructed here. Other proposals embellish Optimality Theory with additional mechanisms: Sympathy Theory (McCarthy 1999a) derives opacity effects by nominating privileged *selector constraints* that pick out 'sympathetic' forms which influence the final outcome; Enriched Input theory (Sprouse, Inkelas and Orgun 2001) derives opacity by admitting *input constraints* that apply not to outputs but to a set of inputs generated from the underlying structure. There may be subtle predictive differences of complexity that can help distinguish between these theories and CCT, but both of them involve an added device which must be stipulated every time there is opacity, offering a less natural account of learning and historical evolution. By contrast, CCT derives opacity from the *same* mechanism as other effects, the cumulative order of constraints. Finally, Optimality Theory's antecedent, Rule-based phonology, has always drawn a significant amount of support from opacity effects (Dresher

1981, Idsardi 2000). However, it has been ill-equipped to solve the problem of predicting and constraining opacity, even though subregularities were observed informally (Kenstowicz and Kisseberth 1973, 1977). The solution in CCT draws on the resources of contemporary phonology, obtaining the patterns that are observed through the effect of feature-Faithfulness constraints in the course of derivations.