

# Abstract Scales in Phonology

David Mortensen  
dmort@socrates.berkeley.edu  
*University of California, Berkeley*

June 4, 2004

There is one glory of the sun, and  
another glory of the moon, and  
another glory of the stars: for one  
star differeth from another star in  
glory. So also is the resurrection of  
the dead.

---

1 Corinthians 15:41-42a

## 1 Introduction

Chain shifts have always been a problem for formal phonology, but have never been a problem to formalize. Rule-based phonology captures them simply by ordering rules appropriately; formalizations of chain shifts within Optimality Theory, while not so straightforward, have proved wholly possible (Kirchner 1995, 1996; Gnanadesikan 1997). From a historical standpoint, there is something to be said for the rule-based approach. Many chain shifts seem to have originated as the result of the interaction of two or more sound changes (1), which the speakers of the languages in question have encoded as synchronic phonological relationships (2):

(1) a.  $y > z / A\_B$  (remaining  $y$  elsewhere)

b.  $x > y / A\_B$  (remaining  $x$  elsewhere)

(2) a.  $/y/ \rightarrow [z] / A\_B$

b.  $/x/ \rightarrow [y] / A\_B$

In such a framework, there is no sense in which the two sound changes are related. The encoding of the chain-shift is simply an epiphenomenon of rule ordering. For many chain-shifts, this is historically appropriate: their chain-like nature is an accident of history.

However, the growth of Optimality Theory has concentrated the attention of phonologists upon finding unified explanations, in the form of universal or motivated constraints, for disparate phenomena. This analytic tendency favors accounts of chain-shifts that capture the shifts as unified processes over those that treat such shifts as independent and coincidental facts. In fact, OT has made chain-shifts more interesting in a number of ways. Working in a framework where rule ordering is no longer available as an analytic option for dealing with counter-feeding opacity, OT phonologists have made a number of proposals for formalizing chain-shifts without derivations (Kirchner 1995, 1996; Moreton and Smolensky 2002; Łubowicz 2003). Other scholars have made proposals that, while not geared specifically towards chain shifts, are nevertheless germane to the issue of scalar phenomena of this kind (De Lacy 2001, 2002). These developments have given impetus to the analysis of chain-shifts that do not appear to be accidental interactions of unrelated processes (Gnanadesikan 1997). Take the vowel raising chain-shifts that are found in many languages:

- (3) a.  $a \rightarrow \varepsilon \rightarrow a \rightarrow i$   
 $\text{ɔ} \rightarrow \text{o} \rightarrow \text{u}$  (Nzebi; Clements 1991)
- b.  $a \rightarrow e \rightarrow i$   
 $\text{o} \rightarrow \text{ɔ} \rightarrow \text{u}$  (Shuijiping Hmong; Xian 1990, Niederer 1998:145,148)
- c.  $\text{ɔ} \rightarrow \text{o} \rightarrow \text{u}$  (Jiaotuo [=Zongdi] Hmong; Niederer 1998:140)
- d.  $e \rightarrow i \rightarrow i^y$   
 $\text{o} \rightarrow \text{u} \rightarrow \text{u}^w$  (Etxarri Navarrese Basque; Kirchner 1995)

It seems quite clear that a related cause is behind the raising of all the vowels in such a chain, in a fashion that cannot be properly captured by phonological rules operating on binary features, but which might be captured through the right kind of constraint, assuming it can refer to some grade across which the vowels are distributed (such as relative height, sonority, or duration).

Chain shifts present two major problems to OT, and different analyses have addressed these problems to different degrees and in different ways. The first is the MOVEMENT IN A DIMENSION PROBLEM—some chain shifts, like those in (3), appear to remap inputs to outputs in some direction along a scale. The second issue is the INCREMENT PROBLEM—plain Correspondence Theory (markedness, faithfulness, and no constraint conjunction) can model neutralization towards an endpoint but not stepwise movement towards an endpoint. Both of these problems have been addressed in various ways in the OT literature on chain-shifts and other recent proposals in OT are relevant to one of both of these problems as well.

This paper will argue that both of these problems may be best resolved if there are abstract scales that function as representational meta-data, allowing both input-output and structural harmony constraints to refer to the relative positions of representations along

these scales. After critically examining some of the earlier approaches to chain-shifts and other scalar phenomena in OT, I will show that the existence of meta-representational scales is independently motivated by the existence of other types of scalar phenomena in phonology, including anti-faithfulness “bounce-back” and cline effects. I will show that this formal mechanism can provide an insightful analysis of previously obscure data in tonology of Mong Leng and Western A-Hmao.

## 2 Formalizations of Scales in Optimality Theory

From its theoretical beginnings, Optimality Theory has been concerned with the encoding of scalar relationships. The formalism is, at its very heart, a theory of constraint hierarchies.

### 2.1 Hierarchies of markedness constraints

Hierarchies of “markedness” constraints, in Optimality Theory, naturally produce certain scalar effects, and analyses of scalar phenomena that need to employ only this mechanism are conceptually economical, since the existence of markedness constraints and some harmonic or stochastic ranking of them is assumed in almost all versions of the theory. The earliest form of Optimality Theory, as presented in Prince and Smolensky (1993) attempted to derive the scalar facts of Berber syllabification from a hierarchy of markedness constraints penalizing segments for occurring in specific syllabic positions (peaks and margins). These hierarchies were seen as deriving from universal facts of sonority, and while the hierarchies themselves were not held to be universal, the ordering relationships that held among various members of those hierarchies were seen as universally fixed.

#### 2.1.1 Fixed markedness hierarchies

Markedness Hierarchies of this type provide one solution to the Movement in a Dimension problem: the hierarchies are derived from some natural phonetic—or otherwise substantial—scale, and universally penalize structures commensurate to their position on such a scale. They do nothing, however, to allow a solution to the increment problem. A hierarchy of markedness constraints not dominated by a faithfulness constraint will favor outputs that are at the bottom of this scale—that is, a neutralization will result. On the other hand, a hierarchy of markedness constraints dominated by faithfulness constraints will, other things being equal, have no effect at all. This problem results from the fact that a scale encoded only in a ranking of markedness constraints cannot be referred to by the grammar except in these constraint’s evaluation of candidates. Such constraint hierarchies provide a possible solution to the Movement in a Dimension Problem but give no insight into the Increment Problem.

## 2.2 Local constraint conjunction and gradience

Kirchner (1995, 1996) presented solutions to both the Movement in a Dimension Problem and the Increment Problem. Assuming that the grammar could make direct reference to phonetic facts, Kirchner proposed that certain chain shifts are movements along a phonetic grade with a single motivating constraint. Vowel raising chain shifts, for example, are motivated by a constraint penalizing output vowels gradiently relative to their duration (in effect, forcing them to rise in the vowel space).

In order to deal with the increment problem, Kirchner (1995) employed the notion of distantial faithfulness—constraints requiring outputs to be within a certain perceptual distance of corresponding inputs. Kirchner (1996) presents a more theoretically orthodox implementation of the same idea: “distantial faithfulness” was replaced by a local conjunction of faithfulness constraints. Such a conjoined constraint is violated if and only if both of the constraints that are conjoined are violated. Interacting with gradient constraints of the type advocated by Kirchner, such constraints are able to model chain shifts in a way that makes them appear to be a unified, rather than an accidental, phenomenon. The cost of such an approach is giving the grammar direct access to gradient phonetic facts. This approach does not adapt itself naturally to chain shifts that have been rendered phonetically unnatural by subsequent sound changes but have retained their chain-like character.

## 2.3 Scalar representations

### 2.3.1 Ternary scales

Gnanadesikan (1997) proposes that some distinctions previously analyzed as privative or binary are better captured in terms of ternary scales and constraints that can refer to these scales. There are two important insights encapsulated in this proposal:

1. Chain-shifts often occur across some identifiable dimension (what I have been calling the Movement in a Dimension Problem).
2. It is not uncommon, cross-linguistically, for the middle element in a three-step chain shift to pattern with the first element in the phonologies of some languages and with the last element in the phonologies of other languages.

Gnanadesikan concentrates upon a scale she proposes, the Inherent Voicing Scale, a scale that captures the relationship between voiced obstruents (in the middle), voiceless obstruents (on one hand), and sonorants (on the other). She proposes, however, that there are other ternary scales as well, including the Consonant Stricture Scale (stop, fricative/liquid, vocoid/laryngeal) and the Vowel Height Scale (high, mid, low). Each point on such a scale is labelled by a natural number 1, 2, or 3.

To interact with these scales, Gnanadesikan proposes a novel group of constraints. The constraints RESIST X and STAY Y are faithfulness constraints that militate against the introduction of a scale value X in the output if it is not in the input and the presence of value Y in the input if it is not in the output, respectively. These constraints interact with assimilation constraints IDENTICAL (which requires adjacent segments to have identical

value specifications on some scale) and ADJACENT (which requires adjacent segments to have adjacent values on some scale). Chain-shifts are analyzed with the relative faithfulness constraints IDENT [X scale] (the output may not have moved on the scale from the input) and IDENT-ADJ [X scale] (the output must be adjacent on the scale to the input).

These constraints and features allow Gnanadesikan to solve both the Movement in a Dimension Problem and the Increment Problem within a limited domain. Chain-shifts occur to partially resolve markedness violations when a complete resolution (i.e. a complete neutralization) would violate IDENT-ADJ but an incomplete resolution would violate only the universally lower-ranked IDENT. Unlike constraint conjunction, though, Gnanadesikan's scales and constraints do not represent a general approach to counterfeeding opacity. It is unclear that this system has anything to say about chain-shifts with more than two steps, or chain-shifts that, synchronically speaking, are not within a single dimension (such as vowel height, voicing, or stricture).

### 2.3.2 *n*-ary scales and scale-referring constraints

In another recent proposal regarding scalar features (though not one concerned with chain-shifts specifically), De Lacy (2002) posits that all features are in fact *n*-ary scales, and devises a novel set of mechanisms allowing for fixed scales of markedness (along which grades may be ignored but not reversed) with completely free constraint rankings. In De Lacy's model, for every scale, there is a set of markedness and faithfulness constraints that refer to that scale. These constraints have the special property of referring not to a single point on the scale (as certain of Gnanadesikan's constraints do) but to a contiguous range that must include the most marked element in the scale. Such sets of constraints are required to be complete; that is, there must be a constraint for each point on the scale.

De Lacy divides these scales into two types:

1. Non-prosodic scales that refer to subsegmental properties and are unidirectional.
2. Prominence scales that refer to properties like tone and sonority are marked relative to their prosodic position. These positions are defined in terms of designated terminal elements (DTEs) and non-designated terminal elements (non-DTEs). For a given prosodic scale, markedness is reversed for DTEs relative to non-DTEs.

The system also includes scale referring faithfulness constraints that preferentially protect marked features (that is, all such faithfulness constraints refer to a range including the most marked end of the scale).

De Lacy demonstrates the usefulness of this formalism for modelling a number of scalar phonological phenomena, including sonority driven stress systems and patterns of place-of-articulation neutralization. However, the system of scales and constraints proposed by De Lacy do not help answer either the Increment Problem or the Movement in a Dimension problem (De Lacy 2002:275fn. 154), nor are they intended too<sup>1</sup>. In part, this is because—unlike Gnanadesikan's constraints—De Lacy's scale referring constraints do not refer to a value on a scale. In other words, De Lacy's scale referring constraints are scale referring

<sup>1</sup>In fact, he expresses skepticism as to whether chain-shifts are scalar phenomena at all.

only in the sense that the existence of a certain scale implies the existence of a set of markedness and faithfulness constraints with specific properties. As this paper will show, De Lacy's position that scales are not reversible is probably too strong.

## 2.4 Contrast preservation

In contrast to De Lacy, who presents a theory of scales that does not address chain-shifts, Łubowicz (2003) presents an analysis of chain shifts that is not overtly scalar. Working in a version of Contrast Preservation Theory (CPT), Łubowicz proposes that chain shifts are a response to the pressures of markedness, on the one hand, and the need to maintain lexical contrasts on the other. The formalism in which this model is stated is a variant of OT that evaluates scenarios rather than individual input-output form pairs. The grammar produces a mapping that maximizes faithfulness, minimalized markedness, and maintains contrasts. Under such a theory, the characteristic I have called the Increment Problem falls out as a consequence of constraints that demand the preservation of contrast (PC constraints). Such a theory does not need to refer to scales in order to maintain increments, because the incrementalism emerges from the tendency for forms which are underlyingly different to be different on the surface.

Łubowicz's formulation of this idea generates a number of interesting predictions about the possible structure of chain shifts. It predicts, for example, that there should be no circular chain shifts, or sets of phonological interactions with more than one point of neutralization. These predictions are evaluated critically in §5.1.3.

## 3 Types of Scalar Phenomena in Phonology

From the standpoint of OT, all phonology is scalar in some sense. For each successful candidate, there are a number of unsuccessful candidates that might be placed on a scale of relative well-formedness for a given constraint ranking. However, there are some phenomena in phonology which are scalar in a more theory neutral sense. These phenomena all suggest that the grammar is aware, in some sense, of relationships among forms that are not simply binary in character. The most obvious of these are chain shifts, but there are other, less known, effects that argue for scales in phonology. These include a phenomenon called "bounce-back", an anti-identity phenomenon in which the end-point of a chain-shift or neutralization is mapped to the penultimate representation on a scale, and cline effects, where the grammar seems to enforce the existence of "slopes" in some dimension across a constituent.

A rather extensive and useful survey of scalar phonological phenomena is to be found in De Lacy (2002). There, De Lacy concentrates upon scalar phenomena that can be described in terms of markedness scales, and makes the strong prediction that, to the extent that "prominence scales" exist, they are not reversible. That is, the grammar of a particular language may ignore distinctions along a scale, but may not reverse them. Contrary to this prediction, however, it appears that there are scalar phenomena that do not flow naturally

from recognized markedness scales or, to the extent that they do, display “markedness-reversals.”

### 3.1 Chain shifts

Chain-shifts are, in their broadest characterization, simply instances of counterfeeding opacity in either synchronic or diachronic phonology. Some such mappings are not scalar in any obvious sense:

(4) a.  $igl \rightarrow igl \rightarrow ig\lambda$  (Barrow Inupiaq; Moreton and Smolensky 2002)

b.  $ktn \rightarrow kn \rightarrow kj$  (Hellendoorn Dutch; van Oostendorp 2002)

There is little reason to believe that the historical processes that give rise to chains of this type are related at all. Two unrelated sound changes, the second of which coincidentally recreates some structure which had been obliterated by the first, can very easily give rise to scenarios of this sort. Furthermore, such chains do not show obvious progress in any dimension.

However, there is a large body of synchronic chain shifts that are apparently scalar in character. Some chain shifts appear to move in the direction of quantity reduction (5):

(5) a.  $V:: \rightarrow V: \rightarrow V$  (Karak; Bright 1957)

b.  $V_1V_2\# \rightarrow V_1\# \rightarrow \#$  (Hidatsa; Kenstowicz and Kisseberth 1977:178f)

Other chains are stepwise assimilations towards some endpoint (6). These include common types of voicing assimilation before nasals and height assimilations triggered by high vowel affixes:

(6) a.  $nt \rightarrow nd \rightarrow nn$  (Nzema; Kirchner 1996)

b.  $mp \rightarrow mb \rightarrow m$  (Mwera; Kenstowicz and Kisseberth 1977:157)

c.  $a \rightarrow \varepsilon \rightarrow a \rightarrow i$   
 $\text{ɔ} \rightarrow o \rightarrow u$  (Nzebi; Clements 1991)

Still other chain-shifts are associated with morphophonological processes in a language, but are not assimilations to some overt target (7). Such mutations include eclipsis (in languages such as Irish), chain tone-lowering in Western A-Hmao, and sandhi-tone conditioned vowel chain shifts in the Mashan group of Western Hmong languages:

(7) a.  $p \rightarrow b \rightarrow m$  (Irish; Ní Chiosáin 1991)  
 $p' \rightarrow b' \rightarrow m'$   
 $t \rightarrow d \rightarrow n$   
 $t' \rightarrow d' \rightarrow n'$   
 $k \rightarrow g \rightarrow \eta$   
 $k' \rightarrow g' \rightarrow \eta'$

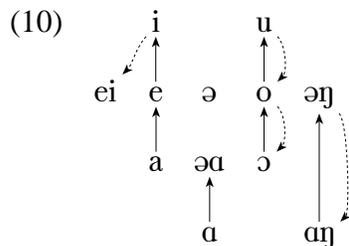
b.  $\uparrow H \rightarrow H \rightarrow M$  (A-Hmao; Johnson 1999)

c.  $a \rightarrow e \rightarrow i$   
 $\text{ɔ} \rightarrow \text{o} \rightarrow \text{u}$   
 $u \rightarrow \text{o} \rightarrow \text{ɔ}$  (Shuijiping Hmong; Xian 1990, Niederer 1998:145,148)

The case of Shuijiping Hmong (7c) is interesting. This set of alternations occurs in the same environments (phonological and morphological) that trigger a tone change (to the low breathy tone—a tone that only occurs in derived contexts). The nature of these alternations depends upon the underlying tone of the affected syllable. If the underlying tone is falling (HM), then a back vowel is raised one increment; if the underlying tone is rising (LM), then a back vowel is lowered one increment (Xian 1990):

- (8) a.  $\text{ʔei}^M + \text{tɕɔ}^{\text{HL}} \rightarrow \text{ʔei}^M \text{tɕo}^L$   
 ‘one’ ‘CLF’ ‘one CLF (road)’  
 b.  $\text{ʔei}^M + \text{tɔ}^{\text{HL}} \rightarrow \text{ʔei}^M \text{t}u^L$   
 ‘one’ ‘step’ ‘one step’
- (9) a.  $\text{toŋ}^M + \text{z}u^{\text{LM}} \rightarrow \text{toŋ}^M \text{z}o^L$   
 ‘NMLZ’ ‘work’ ‘hired farm laborer’  
 b.  $\text{hwei}^M + \text{ŋt}o^{\text{LM}} \rightarrow \text{hwei}^M \text{ŋt}ɔ^L$   
 ‘melon’ ‘pattern’ ‘muskmelon’

These changes are part of a larger complex of changes, shown in (10). The solid lines represent changes that occur when HL-toned syllables become L-toned, while the dashed lines represent changes that occur when LM-toned syllables become L-toned.



These data illustrate two interesting phenomena. The chain shift  $\text{ɔ} \rightarrow \text{o} \rightarrow \text{u}$  is exactly reversed (as  $u \rightarrow \text{o} \rightarrow \text{ɔ}$ ). This suggests that a similar scale is involved in both processes, with one process constituting “ascent” along the scale and the other process “descending” it. Oddly, the factor that determines whether raising or lowering occurs is the underlying tone of the affected syllable, yet this tonal contrast is completely neutralized on the surface.

### 3.2 Bounce-back

Scalar effects may also be manifest when neutralizations and chain-shifts interact with anti-identity requirements. Certain morphological processes require their outputs to be different from their inputs, without specifying the exact way in which the input and output will be different. Take the example of plural stem formation in Jamul Tiipay (a Yuman language of the Kumeyaay group). The most common means of forming plural stems in Jamul Tiipay is by changing the length of the stem vowel (Miller 2001:105):

- (11) a. miiwan → miiwaan  
           lazy        lazy.PL
- b. nyay → nyaay  
           hunt     hunt.PL
- c. nya'war → nya'waar  
           hungry    hungry.PL
- d. u'is     → u'iis  
           sneeze    sneeze.PL
- e. saaw → saw  
           eat        eat.PL

Jamul Tiipay only allows two degrees of length. This means that a short vowel must become long in a plural stem, and a long vowel must become short.

- (12)  $V \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} V$ :

But what happens when this type of anti-identity requirement motivates change in a less constrained space? In compounds in Western A-Hmao (a Western-Hmongic language of the Diandongbei group), there are a set of tone sandhi alternations that occur after the contour tones LM (rising) and HM (falling). In this environment, the tonal alternation generally result in outputs closer to M than the input tone. L↓ (lowest) and L (low) both become M. However, and anti-identity requirement also causes input M to mapped to some other tone (in fact, input L) in the output (Johnson 1999:241):

- (13)  $L\downarrow \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} L \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} M$

- (14) a. ndfiu<sup>HM</sup> + dzfie<sup>M</sup> → [ndfiu<sup>HM</sup>dzfie<sup>L</sup>]  
           sky/season    cold        'autumn'
- b. ?au<sup>HM</sup> + nfiaw<sup>L</sup> → [?au<sup>HM</sup>nfiaw<sup>M</sup>]  
           water     rain        'rain water'
- c. ndfli<sup>LM</sup> + ndfilau<sup>L↓</sup> → [ndfli<sup>LM</sup>ndfilau<sup>M</sup>]  
           rice        glutinous    'glutinous rice'

Such a system of alternations seems to suggest a scalar relationship between the three representations. M is at the “bottom” of this scale, as the endpoint of the neutralization. When bounce-back occurs, it maps M to L as opposed to L↓, suggesting that L is next closest to the bottom—the next best thing to being the “unmarked” mid-tone. The existence of this type of effect provides another reason for recognizing the existence of some type of scales within phonology<sup>2</sup>.

### 3.3 Cline effects

Languages may enforce contours, through some dimension, across a sequence of constituents. Perhaps the most familiar generalization of this sort is that captured by the “sonority hierarchy”, namely that there is an ordering principle such that segments in the onset of a syllable may not be followed by segments of lesser sonority, and that segments in the rhyme of a syllable may not be followed by segments of greater sonority. However, “sonority” as a phonotactic principle is only one type of cline effect. There are other effects within language that display a similar tendencies in the syntagmatic relationships between phonological constituents.

#### 3.3.1 Jingpho coordinate compounds

Some of the clearest of these cases occur in languages with coordinate compounds—compounds displaying no morphological or semantic subordination of one element to the other. In such compounds, the ordering of elements is often predictable on the basis of phonological factors. A nice illustration of this fact is found in Jingpho. The root vowel of the first element in the compound must be “higher” than the root vowel of the second element (Dai 1990:213ff).

(15)

<i>First element</i>	<i>Second element</i>				
i	i	u	e	o	a
u	i	u	e	o	a
e	—	—	e	o	a
o	—	—	e	o	a
a	—	—	—	—	a

- (16) a. *phún*, *kówá* → *phún-kówá*, \**kówá-phún*  
 ‘tree’ ‘bamboo’ ‘trees and bamboo’
- b. *lègo*, *lèta* → *lègo-lèta*, \**lèta-lègo*  
 ‘hand’ ‘foot’ ‘hands and feet’

Such an effect could be described in terms of any number of different scales: vowel height, sonority, or even duration.

<sup>2</sup>Of course, non-scalar analyses of bounce-back effects are possible (just as non-scalar analyses of chain shifts are possible). However, it is probably most intuitive to understand these interactions in terms of a notion of “distance” that is most easily encoded in scales.

### 3.3.2 Mong Leng coordinate compounds

Similarly, the order of elements within Mong Leng coordinate compounds can be predicted based upon the tones of their component parts. A summary of the attested possibilities is given in (17):

(17)

<i>First element</i>	<i>Second element</i>						
H	H	HL	L	L?	MH	ML	M
HL	H	HL	L	L?	MH	ML	M
L	—	—	L	L?	MH	ML	M
L?	—	—	L	L?	MH	ML	M
MH	—	—	—	—	MH	ML	M
ML	—	—	—	—	MH	ML	M
M	—	—	—	—	—	ML	M

- (18) a. tai<sup>L</sup> + tla<sup>MH</sup> → tai<sup>L</sup>tla<sup>MH</sup>, \*tla<sup>MH</sup>tai<sup>L</sup>  
           ‘dish’ ‘spoon’ ‘eating utensils’
- b. tla<sup>MH</sup> + tɕau<sup>ML</sup> → tla<sup>MH</sup>tɕau<sup>ML</sup>, \*tɕau<sup>ML</sup>tla<sup>MH</sup>  
           ‘spoon’ ‘fork’ ‘silverware’

At first glance, it appears that this scale is based upon the height of the first tonal target in each syllable being evaluated:

- (19) H > L > M

However, this view of the subject misses an important synchronic generalization which, it turns out, hinges upon a historical generalization. Note that ML appears to pattern with MH in one dimension, and M in the other dimension. This is due to the fact that Mong Leng ML represents the merger of two historical tone categories.

Four tones may be reconstructed for Proto-Hmongic (and Proto-Hmong-Mien), and each of these tones later underwent a bipartition (based upon the voicing of the onset, which was subsequently neutralized in most of these languages) to yield a system of eight tones. If we group the words in Mong Leng coordinate compounds by their historical tone categories (A, B, C, and D), we can set up the following hierarchy:

- (20) \*A > \*D > \*B > \*C

This hierarchy, translated into the tones of modern Mong Leng, would have the structure

- (21) H, HL > L, L? > MH, ML<sub>1</sub> > M, ML<sub>2</sub>

where ML<sub>1</sub> and ML<sub>2</sub> are ML tones reflecting \*B and \*C, respectively. This same historical hierarchy of tones properly predicts the ordering of coordinate compounds throughout much of the Hmongic family (taking into account the fact that complete mergers have rendered these categories opaque in certain languages of this group). However, the reflexes of these historical tone categories differ to a remarkable degree even among the languages most closely related to Mong Leng (Niederer 1998:249):

	A		D		B		C	
	1	2	1	2	1	2	1	2
(22) Dananshan	HM	ML	M	MH	H↑	L	H	LM
Xuyong	HM	ML	M	LM	HL	Lfi	H	Mfi
Mong Leng	H	HL	L	L?	MH	ML	M	ML
Hmong Lenh	H	ML	MH	ML?	M↓	MLfi	M↑	Lfi
Dashanjiao	H	ML	M	Lfi	HM	Lfi	MH	MHh

These differences are so great that it is probably not possible to devise a synchronic principle that accounts for the tonal constraints on coordinate compounds in more than a few members of the family at one time. In other words, it appears that the learners of these languages have successfully learned, as scalar relationships, phonological patterns with no synchronic phonetic correlates. This phenomenon will be taken up again, and given a more formal analysis, in §5.2.

### 3.3.3 Eastern A-Hmao nominal reduplication

Certain assimilatory and dissimilatory processes have scalar aspects and can be seen as cline effects. An example of such an effect is the vowel disharmony that can be seen in Eastern A-Hmao nominal reduplication (Wang and Wang 1996). This reduplication construction, the rhyme of the (monosyllabic) root of the reduplicant is always either *-u* or *-i*. The quality of this vowel is a function of the quality of the corresponding vowel/diphthong in the base. For some base vowels, the reduplicant vowel must be *i*; for others, *u*; and for still others, both *i* and *u* are acceptable variants:

$$(23) \overbrace{\text{i a ə u i e ai au}}^{\text{u}} \quad \overbrace{\text{y o aw ey}}^{\text{i} \sim \text{u}} \quad \overbrace{\text{u}}^{\text{i}}$$

$$(24) \text{ a. } \text{a}^{\text{H}}\text{ma}^{\text{HM}} \rightarrow \text{a}^{\text{H}}\text{mu}^{\text{H}}\text{a}^{\text{ML}}\text{ma}^{\text{ML}}$$

‘eye’                      ‘eyes, ears, mouth, and nose’

$$\text{ b. } \text{pi}^{\text{H}}\text{ndzau}^{\text{HM}} \rightarrow \text{pi}^{\text{H}}\text{ndzu}^{\text{H}}\text{pi}^{\text{ML}}\text{ndzau}^{\text{ML}}, \quad \text{pi}^{\text{H}}\text{ndzi}^{\text{H}}\text{pi}^{\text{ML}}\text{ndzau}^{\text{ML}}$$

‘demon’                      ‘spirits of every description’ ‘id.’

$$\text{ c. } \text{a}^{\text{H}}\text{ndu}^{\text{HM}} \rightarrow \text{a}^{\text{H}}\text{ndi}^{\text{H}}\text{a}^{\text{ML}}\text{ndu}^{\text{ML}}$$

‘side’                      ‘thereabouts’

Such an effect can be seen most clearly as a scale, along which the rhyme in the reduplicant and the rhyme in the base must differ. The members of the first set share a position on the scale with *i*, and thus must take *u* in the reduplicant. The members of the second set (consisting only of *u*) can satisfy this requirement only by reduplicating with *i*. For the intermediate set, however, either *i* or *u* are always at a different level on the scale, and both options are therefore acceptable.

This relationship is not simple to capture in terms of features, since it seems to encode the relative distance of rhymes from *u* in three different dimensions. The first set includes items that contain no rounded element; the second set includes items that do include a rounded element, but are not *u*; the third set includes items that are *u*. We cannot describe

the relation simply as dissimilation of [round], since *o* and *y* bear this feature just as much as *u* does. We must probably view it, instead, as the result of a mandate that the cline between root vowels in the reduplication construction not be level on the scale given in (23).

## 4 Scales as Representational Meta-data

I propose that these, and a whole range of related data, can be insightfully analyzed in terms of abstract phonological scales that provide the grammar with information about the relationships between representations. This proposal is presented in terms of Optimality Theory, but competes, in explaining empirical phenomena, with a number of widely accepted properties of orthodox OT.

### 4.1 Representations about representations

The abstract scales proposed here are neither metagrammars regulating constraint rankings (like the scales in Prince and Smolensky 1993) nor non-binary representations (like the scales proposed by Gnanadesikan 1997 or De Lacy 2002). Rather, they are a form of representational metadata to which constraints in the grammar may make reference. Such metadata serve to organize possible representations in the grammar into non-exclusive hierarchical relationships.

From the point of view of the grammar, these scales can neither be derived from the representations they rank, nor are they directly derived from some aspect of phonetic substance. That is, they are to be conceived of neither as direct descriptions of some phonetic dimension such as vowel height, the tone space, or the sonority hierarchy (though, for reasons that will be discussed, they will often be nearly identical to such dimensions) nor are they to be seen as hierarchies of markedness. Rather, they are wholly abstract constructs of the grammar of an individual language, which are posited to rationalize phonological patterns within that language.

While the whole utility of these scales comes from the fact that they cannot be derived from facts of representation, they are nevertheless constrained by representational facts, specifically the issue of type. Assume that any phonological constituent can be assigned some type  $t \in T$  (where types, themselves, should be conceived as being structured in an inheritance hierarchy, such that members of some type may also be members of a broader “super-type”, and all members of that “super-type” may in turn be members of a yet more comprehensive type). If scale  $\mathbf{S}$  is a scale of type  $t$ , then all possible representations of type  $t$  are associated with some level on scale  $\mathbf{S}$ . Such a scale could be formalized as a partial ordering on a set (a poset) where the poset is a chain of anti-chains. That is to say, the scale is like an indexed sequence of sets where each set is unordered and contains all of the elements on one level of the scale (informally understood). Here, we will give these scales simpler representation as an indexed  $n$ -tuple of unordered sets:

$$(25) \text{ a. } \mathbf{S} = \langle \{a, b, c\}, \{d, e\}, \dots, \{x, y, z\} \rangle$$

$$\text{b. } \mathbf{S}_1 = \{a, b, c\}, \mathbf{S}_2 = \{d, e\}, \dots, \mathbf{S}_n = \{x, y, z\}$$

In other contexts, it is useful to view such a scale as a function that returns the index of the set in which a representation appears (a natural number between 0 and  $n$ , with 0 representing the top of the scale). If  $S^3$  is treated as a function, then for any two representations  $r$  and  $r'$  of type  $t$ ,  $S(r) = S(r')$  or  $S(r) < S(r')$  or  $S(r) > S(r')$ . That is, any two elements sharing the same type as a scale can always be compared along that scale. We will also assume these scales to be strictly transitive (harmonic). Thus, given the scale  $S$  and the representations  $a$ ,  $b$ , and  $c$ , if  $S(a) < S(b)$  and  $S(b) < S(c)$  then  $S(a) < S(c)$ . Any such scale can be decomposed into a set of pairwise precedence statements of the form  $r > r'$ , but the converse is not true: one cannot compose such a scale out of any possible set of pairwise precedence statements since it is possible to state non-transitive relationships in pairwise terms, and such non-transitive relationships would render the constraints proposed here incoherent.

The grammar may refer to these scales through two families of constraints: input-output constraints, which require a certain scalar relationship between corresponding inputs and outputs, and structural harmony constraints, which penalize certain scalar relationships across the output.

## 4.2 Input-Output constraints: PROD, STOP, and STYMIE

The constraints proposed here encode three different types of input-output relationships: identity, scalar retentiveness, and non-identity. The first two constraint types favor identity between inputs and outputs. STOP is essentially a scalar version of orthodox OT IDENT. It is violated when the members of an input-output pair differ in their position on the relevant scale. Following Kirchner (1995) and Gnanadesikan (1997), I propose that there is also a class of constraints that enforce near-identity (thus solving the Increment Problem in chain shifts). Constraints of this class, here called STYMIE, are violated when members of an input-output pair are not adjacent on the relevant scale. More formal definitions of these two constraint families are given in (26):

(26) a. STOP(**S**)

If **S** is a scale of type  $t$ ,  $\alpha$  is a representation of type  $t$ , and  $\beta$  is an output correspondent of  $\alpha$ , then  $S(\alpha) = S(\beta)$ .

b. STYMIE(**S**)

If **S** is a scale of type  $t$ ,  $\alpha$  is a representation of type  $t$ , and  $\beta$  is an output correspondent of  $\alpha$ , then  $|S(\alpha) - S(\beta)| \leq 1$ .

As second type of constraint is needed to penalize descents on a scale across an input-output pair:

---

<sup>3</sup>Here I will use a boldface capital to refer to a scale as a sequence and the corresponding italic capital to refer to the scale as a function.

(27) RETAIN(**S**)

If **S** is a scale of type  $t$ ,  $\alpha$  is a representation of type  $t$ , and  $\beta$  is an output correspondent of  $\alpha$ , then  $S(\alpha) \leq S(\beta)$ .

Numerous languages also provide evidence of anti-identity effects, and at least some of these seem to be scalar in nature. Such effects, I propose, are motivated by a class of constraints I will call PROD, which penalize input-output pairs where both elements are at the same position on the relevant scale. Such constraints are formally defined as in (28):

(28) PROD(**S**)

If **S** is a scale of type  $t$ ,  $\alpha$  is a representation of type  $t$ , and  $\beta$  is an output correspondent of  $\alpha$ , then  $S(\alpha) \neq S(\beta)$ .

Note that PROD requires that inputs and outputs differ in their positions on the scale. Therefore, an input-output pair where the relevant elements were different, but did not differ in their scalar position, would not satisfy this constraint.

By way of illustration, take the scale  $\mathbf{S} = \langle \{A\}, \{B, C\}, \{D\} \rangle$ . The behavior of these constraints can be seen in the following quasi-tableau:

## (29) The evaluation of scalar IO constraints:

	STOP( <b>S</b> )	RETAIN( <b>S</b> )	STYMIE( <b>S</b> )	PROD( <b>S</b> )
$B \rightarrow B$				*
$B \rightarrow C$				*
$B \rightarrow A$	*	*		
$B \rightarrow D$	*			
$A \rightarrow D$	*		*	
$D \rightarrow A$	*	*	*	

As can be seen, PROD(**S**) is the complement of STOP(**S**), while STYMIE(**S**) and RETAIN(**S**) both mark a different proper subset of the violations marked by STOP(**S**).

### 4.3 Structural Harmony Constraints: \*WAX, \*WANE, \*STAGNATE, and MOST

I propose that output constraints can make reference to scales just as input-output constraints can. These constraints enforce well-formedness conditions out the output in terms of scales. Such constraints can be used to model a variety of common phonological processes and conditions including assimilation, dissimilation, harmony, directional spreading, and positional neutralizations.

These constraints are formally of two types. The first type are cline constraints, which penalize certain syntagmatic relationships among the scale-values of entities within a certain domain.

(30) a. \*WAX(**S**, **D**)

If **S** is a scale of type  $t$ , and **L** is a list of all elements of type  $t$  ( $r_1, r_2, \dots, r_n$ ) within

domain **D**, then for each pair  $(r_i, r_{i+1})$  in **L** such that  $S(r_i) - S(r_{i+1}) = n > 0$ , mark  $n$  violations.

b. \*WANE(**S**, **D**)

If **S** is a scale of type  $t$ , and **L** is a list of all elements of type  $t$   $(r_1, r_2, \dots, r_n)$  within domain **D**, then for each pair  $(r_i, r_{i+1})$  in **L** such that  $S(r_{i+1}) - S(r_i) = n > 0$ , mark  $n$  violations.

c. \*STAGNATE(**S**, **D**)

If **S** is a scale of type  $t$ , and **L** is a list of all elements of type  $t$   $(r_1, r_2, \dots, r_n)$  within domain **D**, then for each pair  $(r_i, r_{i+1})$  in **L** such that  $S(r_i) = S(r_{i+1})$ , mark one violation.

This set of constraints can motivate assimilation and dissimilation along scales, and also require a sequence of subconstituents of some constituent to have a particular scalar relationship to one another with vis-a-vis some scale. Assuming, again, the scale  $A > B, C > D$ , these constraints mark violations as follows:

(31) Evaluation of syntagmatic structural harmony constraints:

	*WAX( <b>S</b> , <b>D</b> )	*WANE( <b>S</b> , <b>D</b> )	*STAGNATE( <b>S</b> , <b>D</b> )
[AA] <sub>D</sub> , [BB] <sub>D</sub> , [BC] <sub>D</sub> , ...			*
[BA] <sub>D</sub> , [CA] <sub>D</sub> , [DC] <sub>D</sub> , ...	*		
[AB] <sub>D</sub> , [AC] <sub>D</sub> , [CD] <sub>D</sub> , ...		*	
[AD] <sub>D</sub>		**	
[AAA] <sub>D</sub> , [BBB] <sub>D</sub> , [BCB] <sub>D</sub> , ...			**
[ABB] <sub>D</sub> , [ABC] <sub>D</sub> , [BCD] <sub>D</sub> , ...		*	*
[ABA] <sub>D</sub> , [CDB] <sub>D</sub> , ...	*	*	
[BBA] <sub>D</sub> , [CBA] <sub>D</sub> , ...	*		*

Any pair of contiguous elements will violate one (and exactly one) of these constraints relative to a specific scale. Thus, these constraints can quite powerfully regulate the types of scalar relationships that can exist between contiguous elements of the same type. Specifically, such constraints can capture a variety of interesting generalizations:

- The tendency for some constituents (onsets, rhymes, feet) to favor a cline in some dimension (such as “sonority” or tone).
- Directional assimilations and spreading.
- Dissimilation and disharmony.

These constraints are related to the assimilation constraints of Gnanadesikan (1997), in that they evaluate scalar relationships between contiguous elements. However, they differ in a number of respects, aside from the nature of the scale to which they refer. In the first place, pairs are only evaluated if they are type-contiguous within some domain:

(32) **Type contiguity:**

Two elements  $\alpha$  and  $\beta$  are type contiguous within domain  $D$  if  $\alpha$  is the  $n$ th element of type  $t$  within domain  $D$  and  $\beta$  is the  $n + 1$ th element of type  $t$  within domain  $D$ .

This notion of type contiguity produces some of the same effects as have been derived in the past via contiguity on autosegmental tiers. That is, some elements are evaluated by the constraint as local even though they are not contiguous within the speech stream.

The second class of output constraints proposed here penalize elements in specific prosodic positions for not being at a terminus of some scale:

(33) **MOST(S, P)**

If  $S$  is a scale of type  $t$  then for each element  $\alpha$  of type  $t$  in prosodic position  $P$ , if there is some  $\beta$  such that  $S(\beta) > S(\alpha)$  and there is no  $\gamma$  such that  $S(\gamma) > S(\beta)$  then mark  $S(\beta) - S(\alpha)$  violations. (Informally, mark a violation for each rung below  $\alpha$  on the scale).

Aside from their atomic structure, constraints in this class are comparable in many ways to the scale referring markedness constraints of De Lacy (2002) (and, therefore, to earlier proposals about scalar markedness such as Prince and Smolensky 1993). Most importantly in this regard, they make reference to particular positions (heads and non-heads of constituents). However, they operate as a single constraint rather than a hierarchy of constraints that is fixed (as proposed by Prince and Smolensky) or freely rerankable (as argued by De Lacy) which counts violations based upon displacement from one end of the scale.

Taking, again, the scale  $A > B, C > D$ , the behavior of this constraint class can be exemplified as follows:

(34) **Evaluation of MOST(S, P)**

	MOST(S, P)
$A_P$	**
$B_P$	*
$C_P$	*
$D_P$	

## 4.4 Theoretical potential

This set of theoretical assumptions would make Optimality Theory far more powerful than an Optimality Theory without any type of representational metadata. This is a significant concern, since Optimality Theory with only standard markedness and faithfulness constraints is already very powerful and, according to proofs presented in Moreton (1999), can model any logically possible mapping except for circular chain shifts and infinite chain shifts. With the addition of abstract scales and constraints that can refer to them in the manner described here, it becomes possible to model certain circular chain shifts. Furthermore, certain patterns that were very difficult to model in classical OT become very simple to model. It is legitimate to fear that such formal power might facilitate analyses that model data without capturing important generalizations.

However, the theory of abstract phonological scales should not be seen as a competitor simply to orthodox optimality theory, but to “paradigmatic mapping” which has been the conventional description of various types of phonological alternations like the tone sandhi patterns in Xiamen Chinese and Taiwanese, a phonetically unnatural circular chain shift (Schuh 1978; Anderson 1978; Moreton 1999; Chen 2000). If we propose that the grammar of a language can employ arbitrary mappings of the type  $ABC \rightarrow DEF$ , then the structure of the grammar imposes no constraints upon the possible types of input-output (and, by extension, output-output) relationships that can exist. At least as importantly, such mappings are not very interesting—they simply reiterate descriptive generalizations, without relating them to any other aspects of the synchrony or diachrony of the language.

Abstract scales allow us to capture generalizations that appear intractable in classical Optimality Theory without resorting to purely arbitrary mappings. These scales should be seen as formal mechanisms by which the grammar captures generalizations about the language that cannot be formalized in terms of the representations themselves. Such scales often appear to have a phonetically natural origin. However, the subsequent changes in the phonetics and the phonological representation of elements related by such a scale do not necessarily erode the reality of the scale within the phonology. The grammar is capable of modelling patterns at a very high level of abstraction, and this is the fact that abstract scales are intended to capture.

#### 4.4.1 A conservative interpretation

One advantage of a formal mechanism such as abstract scales (and the accompanying constraints) is that it allows the phonologist to talk about unnatural relationships while retaining both binarity and concrete (i.e. non-abstract) representations. This is so because representational meta-data allows a decoupling of certain generalizations about the representations from the internal structure of the representations themselves. Such a decoupling makes it unnecessary to encode all of the phonological properties of a segment in the featural representation of the segment. So interpreted, a scalar metadata approach to “crazy rules” would supplement but not displace the formidable body of phonological theory that has been built around contentful phonological representations.

#### 4.4.2 A radical interpretation

The conservative interpretation of abstract scales will be assumed here, but it is necessary to note that a much more radical interpretation is possible and has certain very appealing characteristics (including conceptual economy). Scalar constraints can be used to model many of the same types of phonological patterns that have long been described in terms of distinctive features and feature geometries. It would be possible, in fact, to completely replace all featural complexes with collections of metadata that annotate atomic phonological symbols. The effect of such an analytic choice would be to completely decouple the phonological categories and relationships to which the grammar makes reference from the phonetic correlates of those symbols. Model or none, such relationships exist, but under a model of this type, correlations between phonetic substance and phonological classes are a byproduct of phonetically-conditioned sound changes that first created the alternations

now captured by the grammar in terms of scalar relationships and constraint rankings. That is to say, segments that pattern as a class and share a common phonetic feature do not, synchronically, pattern as a class because they share that phonetic feature, but because they are annotated as members of a particular class (they share a level on some scale). The correlation between phonetic implementation and phonological behavior becomes and epiphenomenon of history—the phonetics of the sounds (inherited by the speakers of a language from their linguistic ancestors) at an earlier stage facilitated some type of systematic misperception that was encoded in the phonology of the language as representational meta-data. Essentially, then, this version of the abstract scale theory would treat “telescoping” alternations and “natural” alternations uniformly.

### 4.4.3 Two factorial typologies

It is conventional, in proposing OT constraints, to take some stock of what possible grammars they imply. The number of constraint classes proposed here is quite large. For eight constraints, a complete factorial typology would describe 256 rankings. This would be neither feasible nor useful here. However, it will be useful to look at the predictions of a few of these constraint classes in interaction with one another.

#### 4.4.3.1 Assimilation

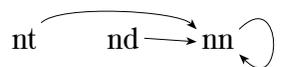
The set of constraints described here can model several types of assimilatory processes, including assimilatory chain shifts. Here, we take the example of nasal assimilation as a help in making concrete the findings of the typology.

Assume the scale  $n > d > t$ , the constraints \*WANE, STYMIE, and MAINTAIN, and the inputs *nt*, *nd*, and *nn*. A summary typology is given in (35):

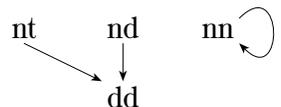
(35) a. MAINTAIN, STYMIE  $\gg$  \*WANE

$nt \longrightarrow nd \longrightarrow nn$   chain-shifting assimilation

b. MAINTAIN, \*WANE  $\gg$  STYMIE

$nt \xrightarrow{\hspace{1cm}} nd \longrightarrow nn$   complete (neutralizing) assimilation

c. STYMIE, \*WANE  $\gg$  MAINTAIN

$nt \quad nd \quad nn$   incomplete (neutralizing) assimilation

If the other two constraints dominate \*WANE, then an assimilatory chain-shift will result (identical to that found in Nzema). If the other two constraints dominate STYMIE, then complete assimilation will occur. If both other constraints dominate MAINTAIN, then an interesting, opaque looking, interaction is predicted to occur. To my knowledge, this type of interaction is not attested between nasals and stops, but is formally similar to patterns known from vowel coalescence where  $/ua, oa/ \rightarrow [oo]$  but  $/aa/ \rightarrow [aa]$  (except that, in these cases, the ordering of the elements is reversed).

## 4.4.3.2 Endpoint interactions

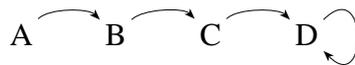
There is also a significant class of interactions that can be modelled in terms of the interaction of MOST constraints, PROD constraints, and STYMIE constraints. These constraints predict four different types of interactions, depending upon their rankings. Assuming a scale  $A > B > C > D$ , the typology is as follows:

(36) a. MOST  $\gg$  STYMIE, PROD



endpoint neutralization

b. STYMIE  $\gg$  MOST  $\gg$  PROD



chain-shift towards endpoint

c. PROD  $\gg$  MOST  $\gg$  STYMIE



endpoint neutralization with bounce-back

d. PROD, STYMIE  $\gg$  MOST



chain-shift towards endpoint with bounce-back

This typology includes two widely attested, and widely discussed, types of interactions: endpoint neutralization and endpoint chain-shift. Additionally, it generates two patterns that have not been widely discussed: neutralization with bounce-back and chain-shift with bounce-back. Neutralizations with bounce-back are attested in both Western A-Hmao (§5.1.6) and Eastern A-Hmao (§5.1.1), and in Jingpho Dai (1990); Lai (2002). No instances of chain-shifts with bounce-back, which are difficult to distinguish from neutralizations with bounce-back on purely formal grounds (unless they involve four or more mappings) have yet been identified as such<sup>4</sup>.

## 5 Scales in Western Hmongic Tonology

The tonal phonologies of Western Hmongic languages provide a nice illustration of the usefulness of abstract scales in accounting for both alternations (tone sandhi) and distributional patterns (the ordering of coordinate compounds according to tone). The tone sandhi cases illustrate the ability of abstract scales to capture, in a meaningful way, historically motivated patterns that have become phonologically unnatural but have remained productive. The coordinate compounding cases illustrate that a set of representations, in a single language, may be organized into multiple orthogonal scales, contrary to the predictions of theories that attempt to derive all scalar effects from markedness hierarchies or scalar representations.

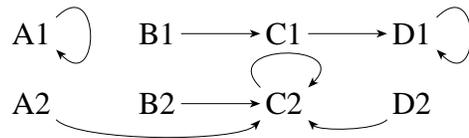
<sup>4</sup>A set of tonal interactions in Eastern A-Hmao could actually be analyzed as a three-step chain shift with bounce-back, but the “neutralization with bounce-back” analysis more insightful both historically and synchronically. Nevertheless, the existence of such similar mappings suggests that chain shifts with bounce-back are possible and will eventually be identified.

## 5.1 Comparative Western Hmongic tone sandhi

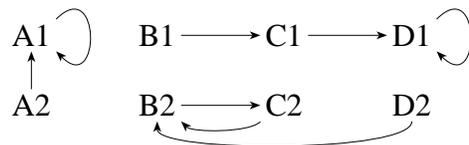
A historical and comparative perspective provides the clearest possible understanding of the nature of Hmongic tone sandhi and why it is problematic for most theories of scales in phonology. Proto-Hmongic had four contrasting tones, typically labelled by specialists as \*A, \*B, \*C, and \*D (Chang 1953; Wang 1994; Ratliff 1992; Niederer 1998). In Western Hmong (and, in fact, in the great majority of Hmongic languages) these four tones have undergone a common Southeast Asian development: they have been bifurcated according to voicing of the onset of the syllable bearing the tone, yielding two “registers” of four tones each (\*A1, \*B1, \*C1, and \*D1 as opposed to \*A2, \*B2, \*C2, and \*D2). The reflexes of these proto-tones show a very high degree of regular correspondence in the daughter-languages of Proto-Western Hmong.

Interestingly enough, the tone sandhi alternations that occur in the “core” group of languages in the Western Hmongic group can also be reconstructed as a function over these historical tonal categories. In these languages, a tone that is a reflex of \*A triggers tone sandhi in the following syllable within a prosodic word. There are a number of strong points of correspondence between the ensuing changes in each of three subgroups of Core Western Hmongic:

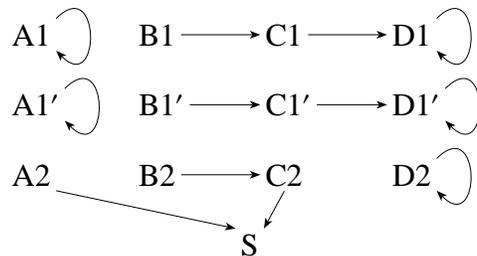
(37) a. Dananshan Hmong (Chuanqiandian subgroup; Wang 1985:20-23)



b. Western A-Hmao (Diandongbei group; Johnson 1999)



c. Shuijiping Hmong (Mashan group; Xian 1990)

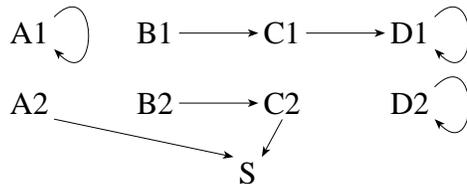


In Dananshan Hmong and Western A-Hmao, the tone changes are all structure preserving, but in Shuijiping, there is a special tone that only occurs in sandhi contexts (here labelled as S). In Shuijiping, the *yin* register (indicated by the number “1”) has undergone a secondary split (conditioned by the aspiration of the syllable onsets), but both sub-registers feature the B1 → C1 → D1 chain that is also present in the Dananshan and Western A-Hmao systems. It is clear that these three systems share too many commonalities to have developed

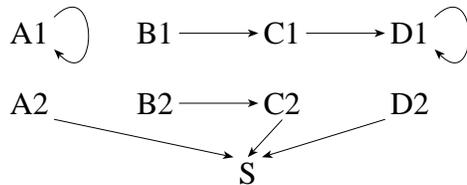
independently, and it is possible to reconstruct a plausible proto-system from which each of the daughter systems could have developed.

The key to accounting for the commonalities in the *yang* register (the series indicated with “2”) is found in the S tone of Shuijiping and the other Mashan dialects, one the one hand. The merger of this sandhi tone with tone B2 in Diandongbei and C2 in Chuanqiandian, can account for the tone sandhi patterns in these dialects. The reconstructed system is given in (38a):

- (38) a. Proto-Core-Western-Hmongic tone sandhi (ancestral to Chuanqiandian, Diandongbei, and Mashan)



- b. Proto-Far-Western-Hmongic tone sandhi (ancestral to Chuanqiandian and Diandongbei)



The members of both the Diandongbei and Chuanqiandian subgroups, which we know on independent ground to be more closely related to one another than to the members of the Mashan subgroup (Johnson 2002), share an innovation that remapped D2, in sandhi, to S. This innovation must have occurred prior to the merger of S with B2 in Diandongbei and C2 in Chuanqiandian. In Chuanqiandian, this merger has led, in essence, to a four way neutralization in sandhi contexts. In Diandongbei, however, it has resulted in an interesting “toggle” between B2 and C2. In Diandongbei, another change occurred prior to the merger of S and B2: A1 was remapped to A2 in sandhi contexts. The fact that these systems, with their different characteristics, can be related so economically suggests that this reconstruction is in fact the right one.

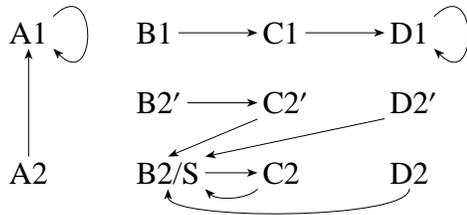
### 5.1.1 Deriving Eastern A-Hmao

This reconstruction allows us to account intelligently even for the related part of the incredibly complicated tone sandhi system of Eastern A-Hmao (as described in detail by Wang and Wang 1984). In this language, the tones B2, C2, and D2 have undergone bifurcation conditioned by the lexical category of the form bearing it (noun versus non-noun)<sup>5</sup>. These

<sup>5</sup>Ratliff (1991) (reference) argues that this tonal split was conditioned by the presence of an affix in the nominal forms. This seems the most likely explanation for this development, given the current state of knowledge.

tones, it seems, were still mapped to S in sandhi context after the split. As in Western A-Hmao, S has merged with B2, yielding a dramatic and involved system of interactions:

(39) Eastern A-Hmao (Diandongbei group)



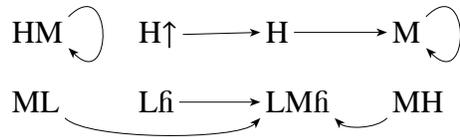
The formal outline of this very involved system suddenly seems coherent when it is compared to the system of alternations that we have already reconstructed for the Diandongbei subgroup (and for Core-Western-Hmongic as a whole). However, the assumptions that we have made in order to win this insight have been slightly different than those implicit in much work on synchronic and diachronic phonology. In the first place, we have treated the symbols manipulated by the phonology as wholly substance free—nothing but abstract categories. We have not discussed, in any way, the phonetic values of the tones or the phonological representations that we have assigned to them. Perhaps more dramatically, we have assumed that, when affected by a split (as in Shuijiping and Eastern A-Hmao) or a merger (as in Dananshan, Western A-Hmao, or Eastern A-Hmao) a tone continues to be treated by the phonology as it was before this process occurred. If, in sandhi context, some other tone was mapped to it, that tone will still be mapped to it. We have assumed that this is natural, despite the fact that such splits and neutralizations must invariably alter the phonological representation of the tone as well as its phonetic realization. These assumptions have allowed us to narrate the story of the development of these alternations, but they have not helped us to explain what the speakers of these languages, at any point in this history, had to know in order to perpetuate these patterns in their speech.

Putting the flesh of phonetic substance upon these abstract skeletons solves some problems, but actually raises others. It turns out that the incredible stability of Proto-Hmong tonal categories, and alternations between them, is matched by an almost as incredible variation in their phonetic implementations across the family, a fact prominently noted by Chang (1953) in his groundbreaking reconstruction of the Hmong-Mien tone categories<sup>6</sup>. Clothed in phonetic descriptions, these tonal alternations appear largely arbitrary<sup>7</sup>:

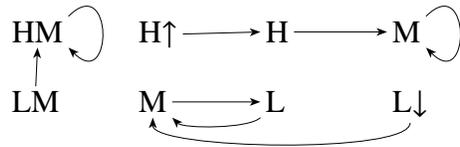
(40) a. Dananshan Hmong (Chuanqiandian subgroup)

<sup>6</sup>Of course, this situation is by no means unique to Hmong, and seems to be a general tendency in the tone systems of East Asia. In this sense, they seem to contrast with families of tone languages like Bantu, which have a relatively high degree of categorial stability (in terms of high toned roots versus low toned roots) but also show a degree of phonetic stability that is quite impressive to an Asianist. It is possible that this difference in relative phonetic stability is a contributor to the differences in the ‘naturalness’ of tonal alternations seen in East and Southeast Asian tone languages as opposed to those in Africa and other parts of the world.

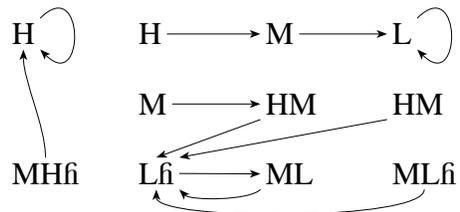
<sup>7</sup>The principle exception to this arbitrariness is the B1 → C1 → D1 chain, which will be discussed below.



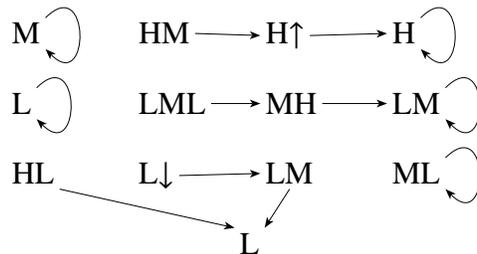
## b. Western A-Hmao (Diandongbei group)



## c. Eastern A-Hmao (Diandongbei group)



## d. Shuijiping Hmong (Mashan group)



The tonal values given here are not necessarily underlying tone values, but rather are phonetic descriptions of the isolation tones for syllables belonging for the equivalent historical category. This is important to note, since syllables from different historical categories with distinct tonological behavior may surface with the same isolation tone. That is, some tonal distinctions are neutralized in isolation but are still referred to in tone sandhi alternations.

### 5.1.2 Western Hmong tone chains and scalar features

One area in which non-abstract representations are helpful is in understanding the B1 → C1 → D1 chain shift which appears, on the basis of data from Far Western Hmongic languages (Dananshan Hmong, Western A-Hmao, Eastern A-Hmao) to have been a lowering chain historically. This can be understood as a side-effect of tonal coarticulation. The ping tones that trigger these alternations seem to have been falling tones historically, and it is easy to imagine the terminus of this fall being realized on in the following syllable, ultimately leading to the perception of step-wise lowering (H↑ might sound more like M in this environment and M might sound more like L). However, we can see from Shuijiping Hmong (which has two reflexes of this historical chain) that progressive lowering is not a *sine qua non* of this chain: neither of the reflexes of this chain are consistently lowering in

their orientation. More cases of this kind can be seen in other languages from the Chuanqidian subgroup, all of them closely related to Dananshan, and having essentially the same set of tone sandhi alternations:

(41) a. MH → M → L (Mong Leng)

b. HM → MH → M (Dashanjiao Hmong; Niederer 1998:107-108)

In original form, as preserved by Dananshan and Western A-Hmao, this chain shift is the kind that could be described reasonably in terms of Gnanadesikan's (1997) ternary scales—movement in a phonological dimension is correlated with movement in a concrete phonetic dimension. However, this type of featural solution is much less desirable, for languages like Dashanjiao, where the chain does not progress through some phonetic dimension (and does not, for that matter, either descend or ascend a “markedness” hierarchy). Analyzing such an alternation in terms of featural scales would imply a very high degree of abstraction.

### 5.1.3 Western Hmongic tone sandhi chains and Contrast Preservation

The interactions here provide a number of interesting counterexamples to the contrast preservation theory of chain shifts advanced by Łubowicz (2003). One interesting prediction of this theory is that chains of phonological mappings should include at most one point of neutralization. That is, if we view mappings as directed pseudographs (directed graphs that allow loops, where nodes can be connected to themselves) then for each connected subgraph, there will be one and only one node with an in-valence greater than one. This hypothesis is suggestive, in the case of Western A-Hmao, where there is a context-free neutralization between tone D1 and tone B2. Under assumptions like those of Łubowicz, we might argue that this neutralization is allowed where others might be banned because it avoids the creation of a mapping with two points of neutralization. However, in Shuijiping, a similar type of neutralization has created just such a mapping. In this language D1' and C2 are both realized as ML in isolation. In sandhi context, C1', D1', and B2 are all realized as LM, but the contrast between C2 and A2 is neutralized—they are both realized as L:

(42) LML → MH → LM → L  
           ↑      ↑  
           L↓   HL

Another prediction made by Łubowicz (2003), but contradicted by these data, is the non-existence of circle shifts. The fact that this prediction falls out of the constraint set seems to be a design feature of the formalism. However, circle chain shifts do occur in tone sandhi (as in the widely discussed cases of tone sandhi in Xiamen, Taiwanese, and other dialects of Southern Min [Chinese]; Wang 1967; Moreton 1999; Chen 2000). This prediction is related to another problematic prediction of Łubowicz's model, namely that chain shifts are always driven by markedness, i.e. that the endpoints of chain shifts are selected on the basis of relative non-markedness. Under such preconceptions, we would apparently have to assume that markedness is a highly language specific notion in order

to properly model these data, even in the cases that are not formally incompatible with Łubowicz's framework. If we did adopt such unconstrained and apparently stipulative notions of markedness, in order to model the alternations, we would still be missing an important generalization. The markedness hierarchies we developed would lead to the preservation of contrasts just in case they were preserved in some earlier diachronic state of the language and to the neutralization of contrasts just in case they were neutralized in some earlier diachronic state. Pretending that the patterns of chain-shift and neutralization that exist in these cases are due to markedness optimization or to contrast preservation seems artificial once the comparative facts are known. One would have to posit wildly different constraints and rankings in order to account for what are, diachronically speaking, the same alternations.

#### 5.1.4 An abstract scalar approach to Western Hmongic tone chains

If these alternations do not emerge from featural representations, from natural markedness scales, or from contrast preservation, then how do speakers encode them in their grammars? Given abstract scales, the problem becomes tractable and these systems, in both their synchronic and diachronic dimensions, can be analyzed insightfully. Such an analysis, while it requires a set of stipulations that are inherently less restrictive than those assumed in more orthodox approaches to scales, is superior both in its historicity its ability to capture synchronic processes. Assume the following, non-crucial, reconstructions of the Proto-Core-Western Hmongic tones:

(43) Proposed values for Proto-Core-Western Hmongic tones<sup>8</sup>:

	A	B	C	D	S
1	*HM	*H↑	*H	*M	
2	*ML	*Lfi	*Mfi	*L?	*L

It should be noted that, at this diachronic stage, there were still voicing distinctions between the onsets of syllables associated with the two registers. However, these voicing distinctions seem to have patterned with tone features rather than with segmental features. Even in languages where this voicing distinction has been lost segmentally, the tonal distinction remains, and any formalization of the tonal alternations must make reference to this category. On this basis, we may propose a scale distinguishing the members of the two registers:

(44) Scale **R** (mnemonic: *Register*)

HM, \*H↑, \*H, \*M > \*ML, \*Lfi, \*Mfi, \*L?

Secondly, there is need for a scale or feature to distinguish between the historically breathy tones, which preferentially alternate with each other, from the modally voiced tones which alternate with other tones and the historically modal/creaky tones which do not participate in any alternations:

<sup>8</sup>Phonetically, B1 may well have been an extra-high tone. It is reflected as such in a number of daughter languages.

(45) Scale **P** (mnemonic: *Phonation*)

Lfi, \*Mfi > \*HM, \*H, \*M, \*ML > \*HM, \*L?

Finally, there is the scale which is descended by the chain-shifts:

(46) Scale **Q** (mnemonic: *Quality*)

H↑, \*Lfi > \*H, Mfi, \*ML > \*HM, \*M, \*L?, \*L

These three scales are sufficient to model the remappings seen in sandhi context<sup>9</sup>the relation between these scales and the tone sandhi patterns can be visualized as follows:

(47) Proto-Core-Western-Hmongic scales and sandhi:

		Q <sub>0</sub>	Q <sub>1</sub>	Q <sub>2</sub>
R <sub>0</sub>	P <sub>0</sub>	*Lfi	→ *Mfi	↘
R <sub>0</sub>	P <sub>1</sub>		*ML	→ *L
R <sub>0</sub>	P <sub>2</sub>			L?
R <sub>1</sub>	P <sub>0</sub>	*H↑	→ *H	→ *M
R <sub>1</sub>	P <sub>1</sub>			*HM

The desired mappings can be predicted with four constraints referring to these scales: STOP(**R**), STYMIE(**Q**), MOST(**Q**), and STOP(**P**). We know, based upon an already establish factorial typology, that STYMIE(**Q**) must dominate MOST(**Q**) (otherwise, chain shifts along **Q** would not exist). STOP(**R**) is not violated, so assume, *ceteris paribus*, that it, too, is undominated. This ranking, combined with the proposed scale, gives the right mappings for some of the tones:

	/MH/	STOP( <b>R</b> )	STYMIE( <b>Q</b> )	MOST( <b>Q</b> )
(48)	(a) H			*
	(b) MH			**!
	(c) M		*!	
	(d) ML	*!		

However, it fails to eliminate some unattested candidates, as in (49):

	/H/	STOP( <b>R</b> )	STYMIE( <b>Q</b> )	MOST( <b>Q</b> )
(49)	(a) M			
	(b) HM			
	(c) MH			*!*
	(d) L	*!		

These candidates are distinguished by the low-ranked constraint, STOP(**P**):

<sup>9</sup>Another scale, distinguishing A1 and A2 (the *ping* tones) from the other tones is necessary in order to actually model the alternations in context. At this point, diachronically speaking, it is likely that these two tones still formed the natural class of falling tones.

	/H/	STOP( <b>R</b> )	STYMIE( <b>Q</b> )	MOST( <b>Q</b> )	STOP( <b>P</b> )
(50)	(a) M				
	(b) HM				*!
	(c) MH			*!*	
	(d) L	*!			

This ranking, in combination with the scales established in (47), will give all the correct mappings for Core-Western-Hmongic. Given this constraint ranking, some aspects of the scales, introduced without justification in (43), become clear. If a tone is not mapped to any other tone in the sandhi environment, then that tone must be at the bottom of the scale referred to by the chain-shift motivating MOST constraint. If there is some tone  $\alpha$  to which no other tone is mapped in sandhi, then there must be at least one scale **S** such that the scale value of  $\alpha$  on **S** is not shared any tone  $\beta$  where  $Q(\beta) > Q(\alpha)$ . In other words,  $\alpha$  needs its space. The register scale (**R**) is motivated on historical grounds independent of the tone sandhi alternations. This scale is needed to get the tones in one register from engaging in mapping interactions with tones in another register. It is not adequate to perform both this function and the function of giving non-target tones “space” on scale **Q**, so the third scale, **P** was proposed, with some historical correlates.

### 5.1.5 From Core-Western-Hmongic to Dananshan

In conventional Optimality Theory, it may be said that the grammar resides almost entirely in a single hierarchy of constraints. In the model we have employed here, the language specific grammar consists of a set of representational metadata as well as a constraint hierarchy. Change in the grammar (e.g. changes in the mappings generated by the grammar) must consist in changes in one of these components. Likewise, grammatical stability reflects the perpetuation of aspects of one or both of these subsystems.

The tonal grammar of Dananshan Hmong, a dialect in the Chuanqiandian subgroup, differs from that of Proto-Core-Western-Hmongic in only a few details, although this might not be apparent on first examination. The tonal alternations attested in Dananshan may be exemplified as follows (Wang 1985:20-21):

(51) a.  $H\uparrow \rightarrow H$

$ntou^{HM} + sa^{H\uparrow} \rightarrow [ntou^{HM}sa^H]$   
 ‘cloth’ ‘blue’ ‘blue cloth’

b.  $H \rightarrow M$

$plou^{HM} + mpua^H \rightarrow [plou^{HM}mpua^M]$   
 ‘hair’ ‘pig’ ‘pig hair’

c.  $ML \rightarrow LM$

$tho\eta^{HM} + tle^{ML} \rightarrow [tho\eta^{HM} + tle^{LM}]$   
 ‘bucket’ ‘water’ ‘water bucket’

d.  $L \rightarrow LM$

qaŋ<sup>HM</sup> + zau<sup>L</sup> → [qaŋ<sup>HM</sup>zɸiau<sup>LM</sup>]  
 ‘bottom’ ‘village’ ‘downhill side of town’

e.  $MH \rightarrow LM$

pe<sup>HM</sup> + ŋkeu<sup>MH</sup> → [pe<sup>HM</sup>ŋgɸieu<sup>LM</sup>]  
 ‘three’ ‘pair’ ‘three pairs’

The whole series of changes between Proto-Core-Western-Hmongic and Dananshan can be modelled by retaining the same constraint rankings, but allowing small, incremental changes in the scales. The principle tonal innovation between PCWHm and PFWHm was the remapping of D2 (L?) to S (L) in sandhi context. This change represents a simple promotion of D2 (L?) along the scale **Q**, resulting in the scales and mappings shown in (52).

(52) Far-Western-Hmongic tone sandhi:

		<b>Q<sub>0</sub></b>	<b>Q<sub>1</sub></b>	<b>Q<sub>2</sub></b>
<b>R<sub>0</sub></b>	<b>P<sub>0</sub></b>	*Lɸi	→ *Mɸi	↘
<b>R<sub>0</sub></b>	<b>P<sub>1</sub></b>		*ML	→ *L
<b>R<sub>0</sub></b>	<b>P<sub>2</sub></b>		*L?	↗
<b>R<sub>1</sub></b>	<b>P<sub>0</sub></b>	*H↑	→ *H	→ *M
<b>R<sub>1</sub></b>	<b>P<sub>1</sub></b>			*HM

But why would such a change occur? This question is probably best addressed in terms of perceptions and inferences. In sandhi context, phonetic perceptual factors lead to the systematic misperception of D2 (L?) as S (L). Listeners thus infer that D2 should alternate with S in sandhi context, and assume, therefore, that they must have misassigned D2 (L?) to **Q<sub>2</sub>**. In “correcting” this non-error, language users posit the new grammar represented in (52).

Another case of misperception must have led to the merger of the sandhi tone S with C2 in the Chuanqiandian languages. It is difficult to determine exactly what the phonetics of either of these tones may have been like at the time this merger took place, but what seems certain is that the (modal voiced) S was mistakenly interpreted by listeners as a special case of the breathy-voiced C2. In response to this mistaken inference, they conflated these two categories, so that mappings to S were redirected to C2, but mappings to C2 were retained. This change entailed the demotion of C2 and B2:

(53) Chuanqiandian tone sandhi:

		<b>Q<sub>0</sub></b>	<b>Q<sub>1</sub></b>	<b>Q<sub>2</sub></b>
<b>R<sub>0</sub></b>	<b>P<sub>0</sub></b>		B2	↘
<b>R<sub>0</sub></b>	<b>P<sub>1</sub></b>		A2	→ C2/S
<b>R<sub>0</sub></b>	<b>P<sub>2</sub></b>		D2	↗
<b>R<sub>1</sub></b>	<b>P<sub>1</sub></b>	B1	→ C1	→ D1

This yields the system that is found in Dananshan dialect and, with allowances for a few dialect specific innovations, the other dialects in the Chuanqiandian subgroup.

### 5.1.6 From Proto-Far-Western-Hmongic to Western A-Hmao

The following are examples of the major tone sandhi alternations attested in Western A-Hmao (Johnson 1999:241):

(54) a.  $H\uparrow \rightarrow H$

$kau^{MH} + ki^{H\uparrow} \rightarrow [kau^{HM}ki^H]$   
 line road ‘custom; Christianity’

b.  $H \rightarrow M$

$dfo^{LM} + mpa^H \rightarrow [dfo^{LM}mpa^M]$   
 oil pig ‘lard’

c.  $M' \rightarrow L$

$ndfu^{MH} + dzfie^M \rightarrow [ndfu^{MH}dzfie^L]$   
 sky/season cold ‘autumn’

d.  $L \rightarrow M$

$?au^{HM} + nfiau^L \rightarrow [?au^{HM}nfiau^M]$   
 water rain ‘rain water’

e.  $L\downarrow \rightarrow M$

$ndfili^{LM} + ndfilau^{L\downarrow} \rightarrow [ndfili^{LM}ndfilau^M]$   
 rice glutinous ‘glutinous rice’

The Western A-Hmao tone sandhi system features two significant innovations beyond Proto-Chuanqiandian. The first is the remapping of A2 to A1 in sandhi context. This change could be reflected, simply, by reassigning A2 to  $\mathbf{R}_1$ .

The second development is more interesting—the merger of S and B2, yielding the interesting tonal toggle found in A-Hmao. There is some evidence that this is merely a surface neutralization. Such neutralizations are attested in other Hmongic languages, such as White Hmong (where the tone B2 is neutralized on the surface with D1, though words with these two historical tones still show the historically expected sandhi behavior; see Ratliff 1992). It is also suggestive that, in Eastern A-Hmao (see 39 on 23), the split of B2, C2, and D2 into two subregisters must have occurred before there was any kind of underlying merger between S and B2, since C2' and D2' are not mapped to B2' (as one would predict in the other ordering of events) but rather to B2, just as in Western A-Hmao (where there was

no subregister split). These facts support the hypothesis that the neutralization between S and B2 occurs as a postlexical phonological process (implying a stratal phonological model where the output of the lexical phonology responsible for mapping C2 and D2 to S is the input of a subsequent phonological grammar, which enforces the mapping of S and B2 to the same output). For WAH, this, analysis would not require any changes in the constraint hierarchy or scale rankings for PFWHm, but only a postlexical phonology that produced the right neutralization mapping (perhaps motivated by featural markedness).

There is a significant problem with this proposal: why, given the surface alternations in A-Hmao, would speakers ever posit a category like S of which they never perceive any direct evidence? It seems likely that speakers would posit such a category only if there was no formal means for modelling the observed set of alternations without positing such a category. However, as we have already shown through a factorial typology, the existence of constraints of the type PROD and STYMIE implies the existence of interactions exactly like that seen in the *yang* register of WAH. Since there is independent evidence for the existence of scalar antifaithfulness constraints of the PROD type, we may assume that the grammar does possess the formal means of modelling this type of interaction without levels. In this mapping, B2 has merged with S:

(55) a. Western A-Hmao tone sandhi (by historical tone category):

		Q <sub>0</sub>	Q <sub>1</sub>	Q <sub>2</sub>
R <sub>0</sub>	P <sub>0</sub>		C2	↔ B2/S
R <sub>0</sub>	P <sub>1</sub>		D2	↗ B2/S
R <sub>1</sub>	P <sub>1</sub>	B1	→ C1	→ D1
R <sub>1</sub>	P <sub>2</sub>		A2	→ A1

b. Western A-Hmao tone sandhi (by surface tone):

		Q <sub>0</sub>	Q <sub>1</sub>	Q <sub>2</sub>
R <sub>0</sub>	P <sub>0</sub>		L	↔ M
R <sub>0</sub>	P <sub>1</sub>		L↓	↗ M
R <sub>1</sub>	P <sub>1</sub>	H↑	→ H	→ M
R <sub>1</sub>	P <sub>2</sub>		HM	→ LM

The mappings shown in WAH imply the existence of a PROD constraint locally conjoined with a MOST constraint. With the currently established ranking, both L (C2) and M (B2) are mapped in output to M (B2):

(56) a.

	/L/ (C2)	STOP(R)	STYMIE(Q)	MOST(Q)
(a)	M (B2)			
(b)	L (C2)			*!
(c)	L↓ (D2)			*!*
(d)	M (D1)	*!		

	/M/ (B2)	STOP( <b>R</b> )	STYMIE( <b>Q</b> )	MOST( <b>Q</b> )
	(a) L (C2)			*!
b.	⊖ (b) M (B2)			
	(c) L↓ (D2)		*!	**
	(d) H (C1)	*!		
	(e) ML (A2)	*!		

Based upon these results, and an already established typology, it is clear that this new conjoined constraint should be ranked above MOST(**Q**) but below STYMIE(**Q**). This ranking now produces the proper mappings for M (and the other tones, as well)<sup>10</sup>:

	/M/ (B2)	STOP( <b>R</b> )	STYMIE( <b>Q</b> )	PROD( <b>Q</b> ) & MOST( <b>R</b> )	MOST( <b>Q</b> )
(57)	☞ (a) L (C2)				*
	(b) M (B2)			*!	
	(c) L↓ (D2)		*!		**
	(d) H (C1)	*!			
	(e) ML (A2)	*!			

### 5.1.7 A-Hmao Tone Sandhi and Markedness

The tonal scale that is attested here in A-Hmao presents an interesting challenge to various putative universals of tonal markedness that have been advanced by, among others, De Lacy (1999, 2001). De Lacy's claim is that there is a universal hierarchy of tonal markedness, one of his prosodic scales, of the form H > M > L (which, he suggests, may simply be a truncation of a longer, but similar, scale). The claim is that there is a universal preference for high tones in DTEs and for low tones in non-DTEs, with mid-tones somewhat disfavored in both environments. The A-Hmao case is interesting because chains in both registers converge on a mid tone. This is not the type of interaction that would follow naturally from assumptions like those made by De Lacy. Interestingly there is some evidence that these syllables which are affected by the tone sandhi processes are prosodic heads. In compounds, for instance, it is the first syllable that is subject to reduction; these alternations occur in the second syllable. When two compounds constitute a coordinate compound, the tonal conditions on the ordering of elements applies to the final TBU in each constituent. These facts tend to suggest that the syllables that are affected by these alternations are a kind of head. It might seem problematic, then, that none of the tonal processes attested here end up raising tones in that position to high. In fact, four of the six mappings attested in this system involve lowering, and it seems quite possible that the whole system represents the phonologization of a depression in pitch following a historical falling tone.

The facts of these alternations are not fatal, at least for a weak form of De Lacy's hypothesis, in which scale non-referring constraints are still allowed to refer to tone. The only prediction that the hypothesis makes is that there are no cases where a DTE will

<sup>10</sup>Following standard assumptions, we assume that constraints that are locally conjoined exist elsewhere in the grammar. Here, we assume that both PROD(**Q**) and MOST(**R**) are very low ranked constraints. However, for the sake of conceptual economy, we will not consider them in our rankings.

receive a low tone just because it is a DTE<sup>11</sup> or a non-DTE will receive a high tone simply because it is a non-DTE. In other words, if the hypothesis is falsifiable, this data cannot falsify it. Instead, they hypothesis simply proves profoundly unhelpful in resolving the issues presented by this data. The strong form of the hypothesis, that *only* constraints referring to a H > M > L scale regulate the distribution of tones, is untenable in light of these and a great deal of other data. Under such a restrictive system, there could be no constraint that would penalize outputs with high tones linked to DTEs that would not also penalize outputs with mid tones linked to DTEs.

De Lacy’s scales, and almost all work in phonology that treats scalar phonology as a markedness effects, assume that, for a given phonological characteristic, there is a single scale with a single ordering of elements or values. This assumption prevents the unnecessary proliferation of “scales”, and has encouraged researchers to explore the versatility of relatively few, putatively universal, scales, such as the sonority hierarchy. Such research has undoubtedly been very valuable, and has elucidated the relationships between a single type of scale and phenomena as diverse as syllabification, vowel reduction, and stress assignment. However, such assumptions have obscured the fact that there are numerous language specific scales which may cross-categorize a single set of elements in several different ways. We will now explore this idea further by looking at an interesting tone-morphology interaction in Hmongic languages.

## 5.2 Cline effects in Mong Leng coordinate compounds

As mentioned in §3.3.2 above, Hmongic languages feature productive tonal phonology that does not regulate alternations. The ordering of elements within coordinate compounds (including the four-word compounds often called *elaborate expressions*) can be predicted on tonal grounds.

- (58) a.  $tu^H + ntshai^L \rightarrow tu^Hntshai^L, *ntshai^Ltu^H$   
           ‘boy’ ‘girl’ ‘children’
- b.  $maj^H + \text{jua}^{MH} \rightarrow maj^H\text{jua}^{MH}, *\text{jua}^{MH}maj^H$   
           ‘lowlander’ ‘Chinese’ ‘foreigners’
- c.  $koŋ^H + mpe^M \rightarrow koŋ^Hmpe^M, *mpe^Mkoŋ^H$   
           ‘fame’ ‘name’ ‘reputation’
- d.  $tai^{ML} + tla^{MH} \rightarrow tai^Ltla^{MH}, *tla^{MH}tai^L$   
           ‘dish’ ‘spoon’ ‘eating utensils’
- e.  $tla^{MH} + \text{t}sau^{ML} \rightarrow tla^{MH}\text{t}sau^{ML}, *\text{t}sau^{ML}tla^{MH}$   
           ‘spoon’ ‘fork’ ‘silverware’

<sup>11</sup>There are languages which seem to contradict this generalization directly. These include Chamorro, in which the stressed syllable in a word attracts a low pitch (Chung 1983). De Lacy (2001) notes, however, that his theory allows for the attraction of low tones to stressed syllables because it does not prevent a language from having a constraint mandating that stressed syllable have tone and a markedness constraint banning high tones everywhere. In other words, the hypothesis advanced by De Lacy makes far fewer predictions than may first appear.

These phonological patterns are also best understood in terms of abstract scales, albeit scales different from those we have used to capture the tone sandhi alternations. The tone system and tone sandhi system of Mong Leng are identical to those of Dananshan Hmong, except that in Mong Leng, tone B2 has merged with tone C2 in all environments (meaning that the mapping of B2 to C2 in sandhi context has been rendered obsolete). The **Q** scale for Mong Leng would look much different from the scale we have already established for talking about coordinate compounds (which we will call **C**):

(59) a. Scale **Q** for Mong Leng:

MH > M, HL, L? > H, L, ML

b. Scale **C** for Mong Leng:

H, HL > L, L? > MH, (ML<sub>1</sub>) > M, ML<sub>2</sub>

This scale, **C**, seems to be of great antiquity. Similar scales seem to be in operation in Hmongic languages outside of the immediate genetic neighborhood of Mong Leng. Somewhat more surprising, the same pattern has been identified in the coordinate compounds of Classical Chinese Ting (1975)<sup>12</sup>. The tones of early Chinese loans in Hmong-Mien languages correlate strongly with their Middle Chinese tone values, with each of the four Middle Chinese tones corresponding to one of the Hmong-Mien tones (Chang 1953; Ratliff 1992; Mortensen 2000). Ting's study demonstrated that over 80% of the coordinate compounds in a corpus of Classical Chinese texts followed the same tonal sequencing principle and hierarchy as the Hmong compounds discussed here:

(60) a. *ping* > *ru* > *shang* > *qu*

b. \*A > \*B > \*C > \*D

This suggests that this pattern was either very “natural” at some point in the past, or more probably, that the pattern was borrowed from one language family into the other, perhaps first as a stylistic tendency which was elaborated into a full fledged principle of the grammar.

Given these hierarchies, it is relatively easy to construct a constraint hierarchy that will give the appropriate orderings. We will need to define an additional constraint that ensures if some tone is mapped to a TBU in the input, a corresponding tone will be mapped to the corresponding TBU in the output.

(61) FAITH-ASSOCIATION(Tone)

For every tone associated with some TBU in the input, the corresponding tone is associated with a corresponding TBU in the output.

---

<sup>12</sup>The pattern in Hmong was discovered independently by the author, following up on suggestions of Johns and Strecker (1987), who subsequently became aware of the pattern in Chinese identified by Ting (1975) when he began investigating the literature on coordinate compounds in Chinese. The fact that both investigators arrived at the same scale independently suggests that the observation it embodies is a robust one.

If we assume a constraint  $*WAX(C)$  that is undominated by  $*WANE(C)$ , we get the following results<sup>13</sup>:

		/tla <sup>MH</sup> /, /tai <sup>L</sup> /	FAITHASSOC(Tone)	$*WAX(C, PrWd)$	$*WANE(C, PrWd)$
(62) a.	☞ (a)	tai <sup>L</sup> tla <sup>MH</sup>			*
	☞ (b)	tla <sup>MH</sup> tai <sup>L</sup>		*!	
	☞ (c)	tai <sup>MH</sup> tla <sup>L</sup>	*!	*	
	☞ (d)	tla <sup>L</sup> tai <sup>MH</sup>	*!		*

		/tla <sup>MH</sup> /, /tau <sup>ML</sup> /	FAITHASSOC(Tone)	$*WAX(C, PrWd)$	$*WANE(C, PrWd)$
b.	☞ (a)	tla <sup>MH</sup> tau <sup>ML</sup>			*
	☞ (b)	tau <sup>ML</sup> tla <sup>MH</sup>		*!	
	☞ (c)	tla <sup>ML</sup> tau <sup>MH</sup>	*!	*	
	☞ (d)	tau <sup>MH</sup> tla <sup>ML</sup>	*!		*

We could also model these types of interactions with an undominated  $MOST(C, \Delta_{PrWd})$  constraint that would block orderings where the “head” (which we are assuming to be the final syllable/foot) of the prosodic word):

		/tla <sup>MH</sup> /, /tai <sup>L</sup> /	FAITHASSOC(Tone)	$MOST(C, \Delta_{PrWd})$
(63) a.	☞ (a)	tai <sup>L</sup> tla <sup>MH</sup>		*
	☞ (b)	tla <sup>MH</sup> tai <sup>L</sup>		**!
	☞ (c)	tai <sup>MH</sup> tla <sup>L</sup>	*!	*
	☞ (d)	tla <sup>L</sup> tai <sup>MH</sup>	*!	**

		/tla <sup>MH</sup> /, /tau <sup>ML</sup> /	FAITHASSOC(Tone)	$MOST(C, \Delta_{PrWd})$
b.	☞ (a)	tla <sup>MH</sup> tau <sup>ML</sup>		
	☞ (b)	tau <sup>ML</sup> tla <sup>MH</sup>		*!
	☞ (c)	tla <sup>ML</sup> tau <sup>MH</sup>	*!	
	☞ (d)	tau <sup>MH</sup> tla <sup>ML</sup>	*!	*

Both of these rankings correctly predict that both logical orderings of words with the same tonal scale value are possible<sup>14</sup>:

		/A1/, /A2/	$*WAX(C, PrWd)$	$*WANE(C, PrWd)$	$*STAGNATE(C, PrWd)$
(64) a.	☞ (a)	A1.A2			*
	☞ (b)	A2.A1			*

		/A1/, /A2/	$MOST(C, \Delta_{PrWd})$	$*STAGNATE(C, PrWd)$
b.	☞ (a)	A1.A2		*
	☞ (b)	A2.A1		*

These two rankings are differentiated, however, with their predictions regarding coordinate compounds with more than three components (which are allowed by some speakers of Mong Leng and a closely related dialect, Hmong Daw):

<sup>13</sup>Note, of course, that these rankings are specific to this construction. In fact, there are a semantically definable set of coordinate compounds in Hmong that do not follow this generalization, and it certainly does not seem to be a general property the morphology of the language. The reader should assume, then, that these rankings either define a cophology specific to this construction (see Inkelas and Orgun 1995; Orgun 1996; Inkelas and Zoll 2000, 2003) or that the constraints are somehow indexed to this construction.

<sup>14</sup>Of course, for individual lexicalized compounds, the order is fixed regardless of this fact. However, both orderings are attested, and either one will be acceptable for novel compounds.

- (65) a. pha<sup>HM</sup> + tai<sup>L</sup> + tla<sup>MH</sup> → pha<sup>HM</sup>tai<sup>L</sup>tla<sup>MH</sup>  
 ‘plate’ ‘dish’ ‘spoon’ ‘eating utensils’

	/pha <sup>HM</sup> /, /tla <sup>MH</sup> /, /tai <sup>L</sup> /	*WAX(C, PrWd)	*WANE(C,) PrWd
b.	(a) pha <sup>HM</sup> tai <sup>L</sup> tla <sup>MH</sup>		**
	(b) pha <sup>HM</sup> tla <sup>MH</sup> tai <sup>L</sup>	*!	*
	(c) tai <sup>L</sup> pha <sup>HM</sup> tla <sup>MH</sup>	*!	*
	(d) tai <sup>L</sup> tla <sup>MH</sup> pha <sup>HM</sup>	*!	*
	(e) tla <sup>MH</sup> pha <sup>HM</sup> tai <sup>L</sup>	*!	*
	(f) tla <sup>MH</sup> tai <sup>L</sup> pha <sup>HM</sup>	*!*	

	/pha <sup>HM</sup> /, /tla <sup>MH</sup> /, /tai <sup>L</sup> /	MOST(C, ΔPrWd)
c.	(a) pha <sup>HM</sup> tai <sup>L</sup> tla <sup>MH</sup>	*
	(b) pha <sup>HM</sup> tla <sup>MH</sup> tai <sup>L</sup>	**!
	(c) tai <sup>L</sup> pha <sup>HM</sup> tla <sup>MH</sup>	*
	(d) tai <sup>L</sup> tla <sup>MH</sup> pha <sup>HM</sup>	**!*
	(e) tla <sup>MH</sup> pha <sup>HM</sup> tai <sup>L</sup>	**!
	(f) tla <sup>MH</sup> tai <sup>L</sup> pha <sup>HM</sup>	**!*

As shown in (65b), the grammar employing \*Wax correctly predicts the one attested pattern, where the “falling” cline holds across all of the contiguous pairs, while the \*Most ranking does not distinguish between the correct output and other candidates. These data support the position that the grammar is not, in evaluating the well-formedness of this type of construction, simply examining one position paradigmatically, and letting the other pieces fall where they may, but is actually enforcing a syntagmatic relationship across the whole construction.

## 6 Conclusions

This paper has argued that grammars know more about the relationships between representations than can reasonably be derived from the representations themselves, and that this information is encoded in a set of abstract scales<sup>15</sup> to which certain ranked but violable constraints can make reference. In other words, it has proposed a model of grammar in which formal operations upon phonological substance are mediated by potentially arbitrary hierarchies. Evidence for this position has been drawn from a variety of phenomena, both synchronic and diachronic. Synchronically, it was argued that such scales are necessary to account for various types of chain shifts and “cline effects” and provide an insightful characterization of bounce-back effects in neutralization. Diachronically, it was show—through an analysis of the development of tone sandhi alternations in Core Western Hmongic—that an abstract scale model allows and incremental understanding of grammatical developments that seem abrupt and arbitrary. Certain other theories of scalar and chain-shifting phenomenon were surveyed, and found not to have these properties, or to undergenerate in significant ways.

<sup>15</sup>Note, of course, that any OT system that includes featural markedness constraints in some ranking is also encoding information about the relationships between representations that are not stored in the representations themselves.

The greatest strength, and perhaps the greatest weakness, of the model proposed here is that it allows “unnatural” alternations to be treated in the same manner as “natural” ones, and predicts that *ceteris paribus*, they should be equally likely. This is a weakness in the sense that it allows the theory to overgenerate badly; this is a strength, however, in that it removes the duplication of explanation that is inherent in many models of phonology, including most versions of Optimality Theory, by allowing diachrony to serve as a filter on possible synchronic states. As such, it may be seen as convergent, in its two aspects (and especially in its radical formulation), with both the Substance Free Phonology of Hale and Reiss (2000) and the Evolutionary Phonology of Blevins (in press). On the one hand, it places the onus of constraining the “naturalness” of alternations purely upon diachronic developments that have independent explanations in terms of the phonetics of perception and the psychology of inference. On the other hand, it assumes that all phonological computations manipulate a uniform set of abstract symbols. This analytic tack makes diachronic phonology relatively more important in the enterprise to understand the sound patterns of the world’s languages. The task of the synchronic phonologist, on the other hand, becomes at once more constrained and better defined.

## Bibliography

- ANDERSON, STEPHEN R. 1978. “Tone features”. In Fromkin (1978), 133–176.
- BLEVINS, JULIETTE in press. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- BRIGHT, WILLIAM 1957. *The Karok Language*. Number 13 in University of California Publications in Linguistics. Berkeley: University of California Press.
- CHANG, KUN 1953. “On the tone system of the Miao–Yao languages”. *Language* 29:374–378.
- CHEN, MATTHEW Y. 2000. *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.
- CHUNG, SANDRA 1983. “Transderivational relationships in Chamorro phonology”. *Language* 59(1):35–66.
- CLEMENTS, GEORGE N. 1991. “Vowel height assimilation in Bantu languages”. In Kathleen Hubbard (ed.), “Proceedings of the Special Session on African Language Structures”, Berkeley: Berkeley Linguistics Society.
- DAI, QINGXIA 1990. *Zangmian Yuzu Yuyan yanjiu [Research on Tibeto–Burman Languages]*. Kunming: Yunnan Nationalities Publisher.
- DE LACY, PAUL 1999. “Tone and prominence”. Ms., Rutgers Optimality Archive [ROA 333-0799].
- 2001. “The interaction of tone and stress in optimality theory”. *Phonology* 19(1):1–32.

- 2002. *The formal expression of markedness*. Ph.D. thesis, University of Massachusetts, Amherst. [ROA 542-0902].
- FROMKIN, VICTORIA A. (ed.) 1978. *Tone: a linguistic survey*. New York: Academic Press.
- GNANADESIKAN, AMALIA E. 1997. *Phonology with Ternary Scales*. Ph.D. thesis, University of Massachusetts, Amherst. [ROA 195-0597].
- HALE, MARK AND CHARLES REISS 2000. “‘substance abuse’ and ‘dysfunctionalism’: Current trends in phonology”. *Linguistic Inquiry* 31(1):157–169.
- INKELAS, SHARON AND ORHAN ORGUN 1995. “Level ordering and economy in the lexical phonology of turkish”. *Language* 71(4):763–793.
- INKELAS, SHARON AND CHERYL ZOLL 2000. “Reduplication as morphological doubling”. Ms., Rutgers Optimality Archive, [ROA 412-0800].
- 2003. “Morphological doubling theory”. Ms. UC Berkely/MIT.
- JOHNS, BRENDA AND DAVID STRECKER 1987. “Lexical and phonological sources of Hmong elaborate expressions”. *Linguistics of the Tibeto-Burman Area* 10(2):106–112.
- JOHNSON, MICHAEL 1999. “Tone and phonation in Western A–Hmao”. *SOAS Working Papers in Linguistics* 9:227–251.
- 2002. “The reconstruction of labial stop-consonant clusters in proto-far-western hmongic”. *Transactions of the Philological Society* 100(1):25–58.
- KENSTOWICZ, MICHAEL AND CHARLES KISSEBERTH 1977. *Topics in Phonological Theory*. New York: Academic Press.
- KIRCHNER, ROBERT 1995. “Going the distance: synchronic chain shifts in Optimality Theory”. Ms., Rutgers Optimality Archive [ROA 66-0000].
- 1996. “Synchronic chain shifts in Optimality Theory”. *Linguistic Inquiry* 27(2):341–350.
- LAI, WIENCE 2002. “Jingpho yu shuang yinjie hecheng ci bian diao de youxuan jieshi [an Optimality Theory analysis of tone sandhi in Jingpho disyllabic compounds]”. Ms., Chinese University of Hong Kong.
- ŁUBOWICZ, ANNA 2003. *Contrast Preservation in Phonological Mappings*. Ph.D. thesis, University of Massachusetts Amherst. [ROA 554-1002].
- MILLER, AMY 2001. *A Grammar of Jamul Tiipay*. Berlin: Mouton de Gruyter.
- MORETON, ELIOT 1999. “Non-computable functions in Optimality Theory”. Ms, University of Massachusetts, Amherst. [ROA 264-1199].

- MORETON, ELLIOT AND PAUL SMOLENSKY 2002. "Typological consequences of local constraint conjunction". In Line Mikkelsen and Christopher Potts (eds.), "WCCFL 21", Somerville, MA: Cascadilla Press.
- MORTENSEN, DAVID 2000. "Sinitic loanwords in two Hmong dialects of Southeast Asia". Unpublished honors thesis.
- NÍ CHIOSÁIN, MÁIRE 1991. *Topics in the phonology of Irish*. Ph.D. thesis, University of Massachusetts, Amherst.
- NIEDERER, BARBARA 1998. *Les langues Hmong–Mjen (Miáo–Yáo)—Phonology Historique*. Number 07 in Lincom Studies in Asian Linguistics. München: Lincom Europa.
- VAN OOSTENDORP, M. 2002. "Morphological recoverability in dialects of Dutch". Ms., Meertens Instituut, Royal Netherlands Academy of Arts and Sciences. <http://www.vanoostendorp.nl/pdf/recoverable.pdf>.
- ORGUN, ORHAN 1996. *Sign-based morphology and phonology: with special attention to Optimality Theory*. Ph.D. thesis, University of California, Berkeley.
- PRINCE, ALAN AND PAUL SMOLENSKY 1993. "Optimality theory: Constraint interaction in generative grammar". Ms., Rutgers Optimality Archive [ROA 537-0802].
- RATLIFF, MARTHA 1991. "The development of nominal/non-nominal class markings by tone in Shimen Hmong". In Laurel Sulton, Christopher Johnson, and Ruth Shields (eds.), "Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society", volume 17. Berkeley Linguistics Society, Berkeley: Berkeley Linguistics Society.
- 1992. *Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes, and Expressive Phrases in White Hmong*. Number 27 in Monograph Series on Southeast Asia. DeKalb Illinois: Northern Illinois University Center for Southeast Asian Studies.
- SCHUH, RUSSEL 1978. "Tone rules". In Fromkin (1978).
- TING, PANG-HSIN 1975. "Tonal relationship between the two constituent of the coordinate construction in the Analects, the Meng-tze, and the Book of Odes". *Bulletin of the Institute of History and Philology, Academia Sinica* 47(1):17–52.
- WANG, FUSHI (ed.) 1985. *Miao–yu Jianzhi*. Zhongguo Shaoshu Minzu Yuyan Jianzhi Congshu. Beijing: Minzu Chubanshe.
- WANG, FUSHI 1994. *Miaoyu Guyin Gouni* [Reconstruction of Proto–Miao Language]. Tokyo: Institute for the Study of the Languages and Cultures of Asia and Africa (ILCAA).

- WANG, FUSHI AND DEGUANG WANG 1984. “Guizhou Weining Miaoyu de shengdiao [the tones of the Miao speech of Weining, Guizhou]”. In Maoji Fu (ed.), “Papers on the Minority Languages of China”, Chengdu, Sichuan: Sichuan Minzu Chubanshe, 91–134.
- 1996. “Guizhou Weining Miaoyu dai qianjia chengfen de shuangyin jie mingci de xingtai bianhua”. *Minzu Yuwen* (1):34–36.
- WANG, WILLIAM S.-Y. 1967. “Phonological features of tone”. *IJAL* 33(2):93–105.
- XIAN, SONGKUI 1990. “Guizhou Ziyun Shuijingping Miaoyu he Wangmo Xinzhai Miaoyu de liandu biandiao [tone sandhi in the Miao speech of Shuijiping, Ziyun and the Miao speech of Xinzhai, Wangmo in Guizhou]”. *Minzu Yuwen* (3):71–77.