To appear in Language and Linguistics.
Nov 2003.
m.yip@ling.ucl.ac.uk

Abstract:

Phonological markedness and allomorph selection in Zahao

Moira Yip
University College London

This paper deals with verb stem alternations involving tone, glottalization, and length in the Chin language Zahao (Osburne 1975), spoken in Burma. The emphasis is on tone, but the length facts are extremely interesting, and are given a preliminary treatment in the final section. It is argued that verbs must list stem alternants in the lexicon, but the possible pairs, and the choice of alternant in a given environment, are controlled by markedness. The data offer a theoretical challenge for an output-based grammar like Optimality Theory (Prince and Smolensky 1993), in which the lexicon plays a relatively minor role in the grammar since it is inaccessible to direct control. The Zahao data adds to the list of well-known cases in which allomorphs must be listed in the lexicon, and allomorph selection is then controlled by phonological markedness (e.g. French *ma/mon, ce/cet*, see Tranel 1998). In Zahao, markedness indirectly constrains the set of possible tonal allomorph paradigms, because positional markedness selects the least marked allomorph in one environment, and pressure to realize the full lexical entry selects the more marked allomorph elsewhere. Output-output constraints limit the permissible difference between the two allomorphs. H and L tones tie on markedness, and exchange rules are analyzed as the result of this markedness parity.

Keywords: tone, length, allomorphy, chainshift, paradigm

Submitted to Language and Linguistics, Jan 2003.

Phonological markedness and allomorph selection in Zahao

Moira Yip
University College London

Introduction[1]

In Optimality Theory (Prince and Smolensky (1993); henceforth abbreviated to OT), markedness is usually viewed as the driving force behind phonological alternations, and the lexicon plays a relatively minor role in the grammar since it is inaccessible to direct control. However, there are well-known cases in which allomorphs must be listed in the lexicon, and allomorph selection is then controlled by phonological markedness (e.g. French *ma/mon, ce/cet*, see Tranel 1998). In this paper, markedness indirectly constrains the set of possible tonal allomorph paradigms, because positional markedness selects the least marked allomorph in one environment, and pressure to realize the full lexical entry selects the more marked allomorph elsewhere. H and L tones tie on markedness, and exchange rules are analyzed as the result of this markedness parity. Chainshifts are the result of intra-paradigm conjoined markedness.

The data involve verb stem alternations involving tone, glottalization, and length in the Chin language Zahao (Laizo), spoken in Burma. The emphasis is on tone, but the length facts are extremely interesting, and given a preliminary treatment in the final section. Although Zahao is a Tibeto-Burman language, tonally and in terms of syllable structure it is not unlike some of the southern Chinese dialects. It thus deserves the attention of anyone interested in Chinese tonal systems.

After some preliminaries, I set out the tonal alternation data in section 3, and demonstrate that secondary stems are never tonally more marked than primary stems. In sections 4 and 5, I draw attention to the chainshifting character of the system, and the circularity of the L ~ H and H ~ L pairings, and summarize previous OT approaches to similar phenomena. In section 6, I

[1] This paper has benefited greatly from the input of two anonymous reviewers, and the comments of audiences at the Tenth Manchester Phonology Meeting, University of Manchester, and the 8th International Symposium on Chinese Languages and Linguistics, Academia Sinica, Taipei, most particularly Laura Downing, Dan Everett, Paul de Lacy, Matthew Gordon, Larry Hyman, Ian Maddieson, and Glyne Piggott. All errors and misjudgements are of course my responsibility.

present the analysis, and section 7 discusses its implications. Section 8 moves on to the length facts, and establishes the main generalizations, but does not offer a formal analysis. The data here and throughout comes from Osburne's (1975) description of Zahao Chin. All references to Osburne throughout are to this work. I will mark tones with Ø, L, H, and LH for rising.

1. Syntactic or prosodic context for stem type

In Chin languages (Tiddim: Ostapirat 1998; Lai: Melnick 1997a,b; Daai: Hartmann-So 1989) it is common for verbs to have two stems. 'Primary stems' are used for main verbs, including the main verb of each half of conjoined sentences. Since these are verb-final languages the verbal complex is usually sentence-final, but the stem itself is often not, since it may have various suffixes. 'Secondary stems' are used in subordinate clauses, which are usually near the start of the sentence. This includes relatives, adjuncts, also causatives, benefactives, and resultatives. In the sentence below, the main verb *suaŋH* 'cook' is in its primary stem form, and the verb *hmuʔ* 'saw' in the relative clause is in its secondary stem form..In the sentences below I mark only the verb stem tones. *hmuʔ* is toneless, and *suaŋH* is high.

(1) nunaaw kahmuʔmi ʔin rool ʔasuaŋH
woman I-see-one(agent) food she-cook
'The woman I saw was cooking food'

However, if a main verb is unfocussed the secondary stem may be used. Depending on the question to which it is an answer, the verb /ʔeey/ *'eat*' in *'I ate the chicken'* may be in primary or secondary form, as shown below:

(2) a. ʔaarsa kaʔeeyLH (primary stem, LH tone)
chicken-meat I-eat
Answer to *'What happened?'*/ *'Did you trash the chicken?'* Focus on '*eat'*
b. ʔaarsa kaʔeeyL (secondary stem, L tone)
Answer to *'Did you eat the fish?'* Focus on *'chicken'*, not '*eat'*

This variation is specific to verbs, so either only verbs can occur in the particular context, or only verbs have two allomorphs. The precise environment is obscure, but interestingly different verb forms are used in very similar environments in the South American language Bora (Weber and Thiesen 2000). It may be syntactic (Lehmann (1996), Chhangte (1989)), but the interaction with focus suggests a possible prosodic conditioning. The nature of the primary/secondary difference also suggests a prosodic influence. The secondary stem appears historically to have been formed from the primary stem by addition of a coronal suffix, and vestiges of this remain in the addition of [t] to some vowel-final stems, and in the coronalization of /ŋ/ to [n]. However, synchronically the changes are much more complex than this. Although the details of the alternations vary from dialect to dialect, they typically involve tone, glottalization, vowel length, and final codas. I shall argue that what links the various changes in Zahao is that secondary stems are shorter than primary stems, and less marked tonally.

2. Lexical variability

There is no question that the primary/secondary facts require some sort of lexical encoding. This is because phonologically near-identical primary stems may have different secondary forms as in (a), and it is not possible to take the secondary stem as basic, because similar secondary stems may

have different primary forms, as in (b):

(3)		*Primary*	*Secondary*	
	a.	hmaan L	hmaan L	'be correct'
		hreen L	hren H	'lock up, close'
	b.	laam LH	laam L	'dance'
		hmaan L	hmaan L	'be correct'
		ŋaan H	ŋaan L	'write'

This is reminiscent of the well-known cases in Chinese dialects where two historical categories have merged, but the difference still shows up in sandhi forms. For example, Chen (2000) cites the Min dialect Quanzhou, which has a 31 citation phrase-final tone. Some 31 words become 55 in non-final sandhi contexts, but others become 11. A reviewer points out that similar mergers have also taken place in some Kuki-Chin languages.

3. Tonal markedness reduction in secondary stems

Although there is a lexical component to this system there are also systematic phonological generalizations that must be captured. The striking observation is that secondary stems are never more marked than primary stems**.** Certain highly marked properties, like rising tones, are never found in secondary stems. Any alternations, such as vowel length, or presence vs. absence of tone, always operate so that the secondary stem has the less marked option - shorter vowels, or fewer tones. A purely lexical account, simply listing the two alternants, would fail to capture this tendency towards unmarkedness, and is thus inadequate. Let us begin with some background on tonal markedness.

Primary stems may be L, H, LH or toneless. [2] Underlyingly toneless syllables surface as H after LH, but as L elsewhere. This is very much what happens to toneless syllables in Mandarin, and I shall assume that these tones are assigned in the phonetics, so that the syllables remain toneless in the output of the phonology. Rising tones are unstable. On short CV syllables, LH becomes H except finally, where the vowel lengthens and the rise survives. Before H, LH becomes L. Finally, sequences of rises are grouped from left-to-right into binary domains, and each binary domain gets a single rise spread over two syllables, L.H. For example, (LH.LH.)(LH.LH.)→ (L.H.)(L.H.) Given that we know that rising tones are the most marked cross-linguistically, and frequently restricted to occurring on long syllables, this is not surprising. See Zhang (2000, 2002a,b) for discussion.

Toneless syllables are all stop-final or end in ʔ, glottalized glides or liquids. The preceding vowel is also glottalized. Final stops are probably glottalized, but this is not quite clear. Zahao has no glottalized nasals (but other Chin dialects such as Lai Chin have them: Melnick 1997a,b). Taken together, these observations mean that all toneless syllables have ʔ, and thus that all

---

[2] A reviewer points out that his fieldwork has discovered a falling tone in the Zahao spoken in Falam. Many of Osburne's H tones are falls in Falam, which has a four-way L, H, LH, HL distinction. We cannot know whether Osburne missed a HL, or whether her informants did not have the falls. The analysis here assumes the latter, but if this is wrong it has obvious implications for the analysis. In particular, the reported Falam L primaries with HL secondaries show an increase in markedness in secondary stems, something that I am claiming does not happen in Osburne's Zahao data.

syllables have tone or ʔ, in other words some sort of laryngeal specification. We may formulate this as an undominated constraint SPECIFYLARYNGEAL: 'Every syllable has at least one laryngeal specification, where the set of laryngeal features includes {L, H, ʔ}'.[3]

This minimal requirement for one laryngeal specification is counter-balanced by an upper limit of two laryngeal specifications, since LH is never found on glottalized syllables, including stop-final syllables. In informal constraint terms, then, the language does not allow three or more laryngeal specifications on one syllable, a constraint we could dub *LARYNG^3. It does not seem likely that this can be attributed to a mora-count difference, since stop-final syllables may contrast long and short vowels (doopH 'jump down', vs. fopH 'suck, smoke'), and even the long-vowelled presumably bi-moraic stop-final syllables like [doop] may not carry LH tones. In the remainder of this paper I will not consider candidates with [LHʔ] on the assumption that *LARYNG^3 is undominated.

With this proviso, we can now proceed to unify the tonal facts by assuming that every tonal specification incurs a violation of the markedness constraint than penalizes tonal feature specifications, *TONE, or *T for short. LH thus incurs two violations, and is automatically the most marked. L and H incur one each, and syllables with ʔ incur no violations since they are toneless, and are thus the least marked on the purely tonal dimension. The strictly tonal markedness hierarchy is as follows, where ≻ means 'more harmonic than' or 'less marked than': Ø ≻ L, H ≻ LH, and this surfaces as ʔ ≻ L, H ≻ LH. Interestingly, Hyman and VanBik (2002b) reach compatible conclusions for another Chin language, Hakha-Lai, in which they show that L ≻ HL ≻LH.

3.1 Markedness reduction

With this background in hand, let us look at the purely laryngeal alternations in verb stems. In all cases the left-hand form is the primary stem, and the right-hand column is the secondary stem. A ~ B denotes a primary stem A and its corresponding secondary stem B. There are two possible step-down points on the markedness hierarchy ʔ ≻ L, H ≻ LH., and both are found: LH becomes H or L, and H or L becomes ʔ. (See section 8 on the addition of [t].)

(4) *Rising tones become level*: LH ~ L or H:

laam LH	laam L	'dance'	taar LH	taar L	'hang(up)'
ree LH	reet H	'insert'	na LH	nat H	'be sick'

(5) *Level tones are lost, and stops or ʔ are inserted:* L, H ~ ʔ

doo L	dot	'admire'	khur L	khurʔ	'shiver'
bal H	balʔ	'destroy'	cat H	caʔ	'cut off'

---

[3] One might also wonder whether tone and ʔ are mutually exclusive (Silverman 1997), given that so many glottalized syllables are toneless. Osburne says (p.83) that all primary stems ending in a glottalized sonorant are toneless, and so are many ending in a stop, but she gives many counter-examples:

(1) caak L	'be eager (prim.)'	thulʔ H	'be full (moon)'	(sec.)
kuat H	'send (prim. and sec.)'	zowʔL	'watch'	(p.53)

This second type of markedness reduction is not found with nasal-final stems. This is because glottalized nasals are impossible in Zahao - high-ranked *[+nasal, +c.g.], or *N? for short - and if tone were deleted SPECIFYLARYNGEAL would require glottalization. If *N? and SPECIFYLARYNGEAL are undominated, nasal final stems will have to remain H (or become L, see later):

(6)

/tanH/	*N?	SPECIFYLARYNGEAL	*T
☞ a. tanH			*
b. tan?	*!		
c. tan		*!	

3.2 Markedness stasis

Not all stem alternations reduce markedness. We also find cases in which markedness is held constant:

(7)

L ~ H			H~ L		
baŋL	baŋH	'ressemble'	tanH	tanL	'chop off'
caak L	caak H	'be eager'	malH	malL	'be few'
dooyL	doy?H	'poison'	fopH	fopL	'suck; smoke'

(8)

L or H unchanged:					
bow H	bow H	'bark'	thlaayL	thlaayL	'weigh'
cam H	cam H	'curse'	man L	man L	'reach'
kuat H	kuat H	'send'	-rukL	-rukL	'be six'

(9)

? unchanged:					
dol?	dol?	'swallow'	ko?	ko?	'call'
dap	da?	'grope for'			

Crucially, there are no cases of LH unchanged, showing that it is completely prohibited in secondary position. On the reasonable assumption that secondary stems are less prominent, this is analogous to vowel reduction in a stress language. Just as in Russian or English in unstressed position the vowel inventory is reduced, and the most marked vowels are eliminated, so in Zahao the most marked tone is not permitted. This can be handled in OT by a positional markedness constraint (Zoll 1997, Alderete 1998), *SECCONTOUR, where SEC is a shorthand for the prosodic environment in which secondary stems occur, and *SECCONTOUR bans contour tones (i.e. rises) in such positions. This will dominate the general faithfulness constraint MAX-T (which bans deletion of tones), stopping LH from surfacing in secondary position, but faithfulness will dominate the more general markedness constraint *CONTOUR, allowing LH to surface in primary position. The grammar is thus *SECCONTOUR >> MAX-T >> *CONTOUR. The tableaux show the verb /laamLH/ first in primary position, and then in secondary position.

(10) Rising toned verb in primary context

/laamLH/$_{Primary}$	*SECCONTOUR	MAX-T	*CONTOUR
☞ a. laamLH			*
b. laamL		*!	

(11) Rising toned verb in secondary context

/laamLH/$_{Secondary}$	*SECCONTOUR	MAX-T	*CONTOUR
a. laamLH	*!		*
☞ b. laamL		*	

3.3 No cases of markedness increase

The last observation in this section is that there are no cases at all in Osburne's data in which the secondary stem is *more* marked tonally than the primary stem:

(12) a. No cases of L or H ~ LH or HL
b. No cases of ʔ ~ L or H

In other Chin languages, things look quite different. In Falam Zahao, as a reviewer points out, L ~ HL is found. In Hakha-Lai, nearly all secondary stems have a LH rising tone, raising the possibility that it is a tonal affix. See Hyman and VanBik (2002a) for details. However, a tonal affix analysis does not look promising for Osburne's data, since the primary/secondary pairs LH ~ L and LH ~ H show tone loss, not addition, and L ~ H and H ~L are both found.

Before proceeding to the analysis, there are two more properties of the system that need to be brought out: the chainshifting character of the system, and its circularity. These characteristics place it firmly in the same category as the Min tone circle, the most famous circular chainshift of all.

4. Chainshifting behaviour

In this section I outline a particular OT approach to chainshifts which uses a device called constraint conjunction. The less technically-minded reader may wish to go straight to section 5.

In chainshifts, each category shifts down one in some sequence, but only a one-step change is permitted. In Zahao, LH ~ H ~ ʔ, but crucially, there are no cases of LH ~ ʔ. In OT, alternations are produced by MARKEDNESS >> FAITHFULNESS. But if markedness pressures are paramount, why don't all tones shift to the least marked member of the system, here [ʔ]? Something must be added to the grammar to stop LH shifting all the way to ʔ.

Kirchner (1996) shows that linear chainshifts can be computed by local conjunction of faithfulness constraints, limiting the number of 'steps' between input and output. A conjoined constraint (C1&C2) is violated if and only if both C1 and C2 are violated in some domain. If each 'step' violates either C1 or C2, then two steps will violate (C1&C2). For example, Nzɛbi shows a vowel shift in which a → ɛ → e → i in verbs before certain tense and aspect affixes. Each step changes only a height feature or ATR, but never both, so Kirchner conjoins IDENT ATR with IDENTLOW and then with IDENTHI, and if these are ranked above the pressure to raise all vowels (here shown by a constraint Kirchner calls RAISING) they will limit the raising to a single step. The

tableau shows the case of a → ɛ. The last two candidates [e] and [i] change both [low] and [ATR], so they violate the conjoined constraint IDENT[LOW]&IDENT[ATR]. Assuming that some degree of raising is required, [ɛ] is therefore the optimal candidate.

(13) IDENT[LOW]&IDENT[ATR], IDENT[HI] & IDENT[ATR] >> RAISING >> IDENT F

/a/	IDENT[LOW]& IDENT[ATR]	IDENT[HI] & IDENT[ATR]	RAISING	IDENT F
a			***!	
☞ ɛ			**	*
e	*!		*	**
i	*!	*		***

In Zahao, deletion of a single tone in the H ~ ʔ cases shows that secondary stems penalize any tone at all. In OT terms, *SECTONE >> MAX-T. Now recall that *SECCONTOUR is undominated, because LH never survives in secondary stems. Violations are resolved by deletion of one of the tone, but tone deletion is limited to a single tone, since LH ~ ʔ is blocked, so the conjoined constraints will be MAX-L&MAX-H, and they must dominate *SECTONE, so that one tone will survive. To make the source of the conjoined tones clear, I have replaced MAX-T in this tableau with its two more specific versions, MAX-L and MAX-H, but I will not normally distinguish between them.

(14) Input paradigm with two-tone difference cannot surface; chainshift effect:

/LH/$_{Sec}$	MAX-L&MAX-H	*SEC CONTOUR	*SECTONE	MAX-L	MAX-H
☞ a. H			*	*	
b. LH		*!	**		
c. ʔ	*!			*	*

5. Circularity

Chainshifts in which two or more entities exchange values are called circular chainshifts, or exchange rules. The Nilotic language Luo changes voicing to form the plural:

(15) got gode '*mountain, sg/pl'* debe depe '*four-gallon can, sg/pl'*

A particularly well-known example comes from the Min dialects, which change tones in non-phrase-final position; the entire system is circular (from Chen 2000):

(16) Taiwanese:

53 → 44
↑ ↓
21 ← 22 ← 24

Zahao has a class of H primary-stem verbs that have L in the secondary stem, and others in which

a primary-stem L has a secondary-stem H:

(17)	tham H	tham L	*'feel, touch'*	than L	than H	*'fill'*
	vaakH	vaakL	*'crawl'*	caak L	caak H	*'be eager'*
	khaay H	khaay L	*'chew'*	khaay L	khay?H	*'wear (as earrings)'*

In other words, H ↔ L. There is no segmental conditioning that completely explains the direction of the change. Note though that this circularity is found only at the place in the markedness hierarchy where we have two options that are equally marked tonally because each has exactly one tone: L or H. Thus neither direction of change increases markedness.

Unlike linear chainshifts, circular ones pose a huge problem for OT. Moreton (2004) shows that circular chainshifts cannot be computed at all in OT. The argument goes as follows:

(18) *Why circular chainshifts are non-computable in OT:*
If A →B, it must be because B is less marked than A.
If B → C, it must be because C is less marked than B.
So C ≻ B ≻ A, where ≻ means 'less marked than'.
Thus a hypothetical circular change back from C → A would be an *increase* in markedness, and thus unmotivated in OT.

This is only true for purely phonological operations, as Moreton shows. If the conditioning environment is morphological (i.e. present only in the input) or prosodic at the phrasal level (i.e. present only in the output), then circular chainshifts are statable. His findings confirm those of Anderson and Browne (1973). The conclusion must be that the circular aspect of these changes in Zahao is lexically controlled.

Let me outline a solution. Moreton's argument depends crucially on the assumption that if A →B the change must be driven by markedness. But in Zahao I have suggested that this is *not* true in exactly the one circular case, when L → H or H → L. Suppose instead that the change is driven by the existence of two lexical allomorphs, each of which is striving for survival. If both are to surface somewhere somehow, it must be in distinct environments, and this gives rise to the alternations. So here we have an instance of a morphologically-driven circular chainshift, and not a counter-example to Moreton's argument.

6. Analysis:

6.1 Conceptual overview so far:

So far, I have been concentrating on the phonological generalizations about this system, and these are summarized below:

(19) *In secondary stems:*

1. Markedness stays the same or decreases, but never increases.
   ʔ ≻ H, L ≻ LH, where ≻ means 'more harmonic than' or 'less marked than'
2. Unlike in primary stems, the *upper* limit on laryngeal markedness is one tonal or laryngeal feature:
   *SecContour >> Max-T >>*Contour
3. As in primary stems, secondaries have a *lower* limit of one of H, L, or ʔ. In nasal-final cases, where ʔ may not be added, tones may therefore not be deleted:
   *Nʔ, SpecifyLaryngeal >> *T.
4. Primary and secondary stems differ in markedness by one step only: LH ~ ʔ not found:
   Max-L&Max-H, *SecContour >> *SecTone>> Max-T

5. Circularity results when the two options are equally marked (H and L), and the pressure for change is morphological, not phonological. Both allomorphs are lexically listed.

It is now time to look more closely at the morpho-lexical issues.

6. 2 Lexical encoding

Some primary stem types map to more than one secondary type, and vice-versa, as already explained in section 2, and this must clearly be lexically encoded in some way. The question is how. I will discuss three options, and argue that the only feasible solution is that the allomorphs are listed for each verb.

In OT, lexical variation is sometimes handled by lexically- determined constraint rankings (see Anttila (2002) for discussion and references). A verb root that showed a H ~ L alternation would have MAX-T >> *SECTONE, but a verb root that showed a H ~ ʔ alternation would have *SECTONE >> MAX-T. This is unworkable for Zahao verbs, because one apparently monomorphemic lexical item may have two syllables that would need different rankings: huatHsuakH ~ huatLsuaʔ 'be naughty, childish' (Osburne:105). The argument is similar to Inkelas, Orgun and Zoll's 1997 argument against using lexical diacritics for voicing alternations in Turkish.

A second possibility, and the one that would be taken in most generative approaches, is to assume that the primary is the base form, and posit different abstract underlying representations (UR's) for primaries whose secondaries differ. For example, a primary H that stays H in the secondary stem might be /H/, but a primary H that becomes L might be /HL/. Since HL never surfaces unchanged in Zahao, there would then be processes that simplify HL to H in the primary, but to L in the secondary. Assuming three laryngeal primitives H, L, ʔ, and a requirement that ʔ be final, there are 3 possible singletons and 4 possible pairs: H, L, ʔ, LH, HL, Lʔ, Hʔ. LL and HH are excluded, I assume, by the Obligatory Contour Principle (OCP; Leben 1973).

There are two problems with this approach. First, the abstractness of the UR's is troublesome. Secondly, and more seriously, even this seven-way contrast in available UR's is not sufficient for all cases. Consider nasal-final stems, where ʔ is not allowed. This leaves four possible UR's: H, L, LH, HL, but we find *five* classes of verbs:

(20)

		Primary	*Secondary*	
/L/	hmaan	L	L	'be correct'
/H/	buŋ	H	H	'spill'
/LH/	laam	LH	L	'dance'
?/HL/	hreen	L	H (hren)	'lock up'
??	buan	H	L	'wrestle'

Similar problems arise with verbs of different segmental compositions.

The third option is to assume that each verb root must list both its allomorphs, but to leave the phonological grammar to select which allomorph to use in each context. It is not necessary or desirable to list the allomorphs as specifically "primary" or secondary". It is not necessary because, as I shall show, the grammar can select the right one for the right context. It is not desirable because it would not then be possible to limit the types of primary/secondary pairings that one finds.[4] For example, there could be a verb that was H in the primary and LH in the secondary, but

[4] As a reviewer points out, it is of course possible that the regularities in stem pairs are simply a historical residue, and that nowadays the speaker just lists them, and the grammar plays

no such verbs exist. If the selection is left to the grammar, of course, the *less* marked allomorph will naturally be chosen in secondary stems. It therefore remains to explain why the *more* marked allomorph is chosen in primary stems. In the next section I shall work out this approach.

6.3 An outline of the analysis:

I shall assume, following Tranel (1998) and Perlmutter (1998), that in the lexicon verb stems may list more than one allomorph. For example, the verb 'dance'; will be listed as /laamLH, laamL/. There will (with one exception) be no need to annotate which stem occurs in which environment. If either of these stems surfaces unchanged, the simple Faithfulness constraints such as MAX and IDENT are satisfied.

In the secondary environment, positional markedness will select the less marked form. However, I will also assume that there is pressure to realize all members of a lexical entry, the full paradigm in fact. I will call this constraint MAX-PARADIGM, and it compares input paradigms and output paradigms (see below for a full formulation). As a result, if the same allomorph surfaces in all environments and the other allomorph never surfaces, MAX-PARADIGM is violated. The consequence is that the more marked allomorph will be selected in the primary environment, where information-carrying potential is high (Harris and Urua 2001). Lastly, if both allomorphs are equally marked, the grammar allows either to be secondary, and so in this one instance the lexical entry must note which allomorph is primary and which is secondary.

The following table shows possible lexical entries consisting of one or two tonal allomorphs. The second column singles out the less marked of the two, and this will be chosen as the secondary variant by the positional markedness constraints such as *SECCONTOUR, *SECTONE. The residue will be realized in primary stem contexts, to satisfy MAX-PARADIGM. If the lexicon contains only one allomorph, it will surface unchanged if it is /H, L, or ʔ/.

(21)

Lexical entry	*Less marked = chosen as secondary*	*Residue = primary*
{LH, H}	H	LH
{LH, L}	L	LH
{H, ʔ}	ʔ	H
{L, ʔ}	ʔ	L
{H, L}	H or L	L or H
{H}	H	H
{L}	L	L
{ʔ}	ʔ	ʔ

In addition to the above, there are two logically possible lexical entries that do not surface unchanged: [5]

---

no role at all. Unlike the analysis proposed here, such a non-grammatical approach would presumably claim that the generalizations would not automatically be extended to loans, but I have no way of knowing if this is true or false.

[5] The absence of surface HL in this language is mysterious, especially since the related Falam dialect has HL . Cross-linguistically, HL is more common than LH, yet Zahao has only LH. I must assume undominated *HL, but the reasons are unclear. It is possible that Osburne failed to notice the HL, but she carefully notes that H is falling pre-pausally or under emphasis, and yet reports no H/HL contrast.

(22)	*Non-occurring:*	*Secondary*	*Primary*
	{LH}	*LH	LH
	{LH, ʔ}	*ʔ	LH

A single /LH/ cannot surface as such in secondary position, because undominated *SECCONTOUR will force a change. A /LH, ʔ/ pairing also cannot surface as expected, because the illicit double-leap from LH to ʔ - the chainshifting effect - is ruled out by high-ranked conjoined MAX-L&MAX-H, as discussed above. This constraint will be refined in the next section.

In addition to producing all the possible laryngeal parings, this analysis also correctly fails to produce the remaining unattested ones, given below:

(23)		
	ʔ ~ H	H ~ LH
	ʔ ~ L	L ~ LH

These unattested patterns are ones in which the secondary stem is the more marked of the pair. The positional markedness based account offered here filters out the more marked choice in secondary position, and thus forces the primary stem to be the more marked.

This proposal also allows an understanding of the circularity problem, by which both H ~ L and L ~ H are found. Circularity here comes from the equal markedness of any single tone, plus the allomorph listing. Given a /H, L/ listing, the phonology can select either H or L for the less marked secondary environment, and the other for the primary environment. Individual verbs have settled on one or other option, apparently fixed for each individual verb, and this must be stated in the lexicon, as in /tan$H_{Primary}$, tanL/. Viewed in this way, the circular change is fully compatible with Moreton's work, since it is morphologically controlled. An interesting implication of this proposal is that that there is no universally fixed ranking of the *H and *L markedness constraints, since here in Zahao they must tie in the ranking. See Hyman (2001a) on H-marked and L-marked languages.

Before I present the details of the analysis, one alternative deserves mention. Hyman and VanBik (2002a) offer an account of Hakha-Lai in which the grammar requires Stem 1 and Stem 2 to be different from each other. For example, Stem 2 is normally rising, but if Stem 1 is itself rising, Stem 2 must also become glottalized. Unaware of their paper, I had spent a lot of time trying to work out such an approach for Zahao, using Anti-Faithfulness (Alderete 2001, Horwood 2001), but I had concluded (and still maintain) that it cannot succeed for Zahao, for two reasons. First, the large class of stems that do not change at all (about one third of stems ending in short vowel plus sonorant, for example) show that change is not necessary. Second, the large class of stems that undergo more than one change show that a single difference between Stem 1 and 2 is not always sufficient. For example, we get both tonal and segmental changes on many stems, as in puanL ~ ponH, or reeLH ~ reetH, or baalL ~ balʔH. See section 8 for discussion.

6.4 Fleshing out the proposal:

The constraints that have been discussed earlier fall into two groups, and some of them need a little more discussion before I proceed to show how the analysis works. In the first group are the general markedness and positional markedness constraints.

(24) *TONE: (=*T) No tones.
*CONTOUR: No contour tones.
*SECTONE: No tones in secondary stems
*SECCONTOUR: No contour tones in secondary stems

As I discussed in section 1, the positional context in question is probably prosodic. The data I have are not enough to pin this down, but one possibility might be to explore the notion of Intonation Phrase-tail from Harris & Urua (2001), in which certain positions have less ability to bear information than others, and thus in these positions contrasts are reduced. I leave this question for future research, but ultimately *SECTONE and *SECCONTOUR can, one hopes, be more perspicuously stated.

The second set of constraints are all faithfulness constraints, some more familiar than others. The last constraint is re-formulated here, for reasons I explain below:

(25) FAITH-INPUTOUTPUT (FAITH-IO): In an output paradigm, each member must be identical to some member of the input paradigm
MAX-PARADIGM: All members of an input paradigm must be realized somewhere in the output paradigm.
MAX_{OO}-L&MAX_{OO}-H: Members of an output paradigm may not differ by both MAX-L and MAX-H.

FAITH-IO compares each paradigm member to one of the input members. It is a shorthand for the set of IO-Faith constraints that includes MAX , MAX-T, IDENT, DEP, and DEP-T. It will be violated by deletion or insertion of tone or ʔ, or by a change in laryngeal features, such as H to L, or L to ʔ.

MAX-PARADIGM compares input and output paradigms. The need to assess complete paradigms as output candidates has been recognized by a number of workers, including Steriade (2000), and especially McCarthy (2002) on intra- paradigm comparisons.[6] Usually, however, the input is a single root or stem that appears in different guises in different members of the paradigm. What is new here is the notion that the input is also a sort of paradigm. In Zahao, I have argued that the primary and secondary stems cannot be derived one from the other, but must both be listed in the lexicon. This listing of two allomorphs constitutes an input paradigm, just like the French first person singular possessive *ma/mon/mes*. In Zahao, the input paradigm for 'dance' is /laamLH, laamL/. If the output has each of these in some context, MAX-PARADIGM is satisfied. It does not care which is primary and which is secondary, just that they both surface. On the other hand if it surfaces as [laamLH] in all contexts, or as [laamL] in all contexts, MAX-PARADIGM will be violated. Note particularly that the input paradigm is an unannotated set. The choice of one as primary and one as secondary is accomplished by positional markedness. This therefore rules out verbs in which the primary surfaces as less marked than the secondary.[7]

The last constraint, MAX_{OO}-L&MAX_{OO}-H is a conjoined constraint that limits intra-paradigm differences to one step, following Kirchner. It is a more refined version of what I earlier (§4) called MAX-L&MAX-H. The refinement is necessary for technical reasons. Since the primary and secondary are each related directly to the input by FAITH-IO, in a hypothetical LH primary and ʔ secondary verb stem each of LH and ʔ is fully faithful to its input, /LH, ʔ/. Since such stems do not exist, what is illicit is an output paradigm in which the two component outputs, LH and ʔ, differ

[6] A reviewer reminds me that paradigm-based models raise difficult issues with respect to lexical processing, frequency effects, and learnability. I agree, and the interested reader might consult Reiss, Hale and Kissock (1997), and Adam (2002).

[7] A reviewer suggests that /L/ could be the default in Stem 2. However, this fails to explain the fact that both L~H and H~L pairs seem to have about the same frequency (in sonorant-final stems). I will not pursue this suggestion further here.

by two steps. The relationship is thus not an IO relationship, but one between two outputs, or output-output faithfulness, FAITH-OO, specifically MAX_{OO}-L&MAX_{OO}-H. On the need for output-output faithfulness relationships, see Benua (1997), McCarthy (2002), and many others.

In the remainder of this section I work through the complete set of lexical entries given in (21). In each case the input is a set or list of either one or two allomorphs, and the output candidates are the primary and secondary realizations, in that order. The secondary stem is in italics. I consider all the candidates that use unchanged input allomorphs as outputs, or that reduce them in markedness.

Consider tableau (26). Let me start with the fact that LH is never found in secondary stems. This is the result of high-ranked *SECCONTOUR, which rules out not only candidate (26c), but also (26d), in which both the LH and the H input allomorphs do surface, but with the LH in the secondary stem. Despite its markedness, the LH must surface somewhere, because of MAX-PARADIGM, which rules out the least marked candidate (b). Thus (a) is the actual winner. The tableau for a /LH, L/ stem would be identical.

(26) /LH/ can only be realized in the primary stem, and MUST be realized there:

{LH, H}	*SECCONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
☞ a. LH, *H*				*
b. H, *H*		*!		
c. LH, *LH*	*!			* *
d. H, *LH*	*!			*

Now consider stems in which both allomorphs have simple level tones, shown in (27). MAXPARADIGM requires both the H and the L to surface, but doesn't care where. Thus candidates (a) and (b) will tie, and the difference between a L ~ H and a H ~L verb must be lexically marked.[8]

(27) /H, L/ gives two equally valid outputs, because L, H tie on markedness. The two must then be lexically annotated.

{H, L}	*SECCONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
☞ a. H, *L*				
☞ b. L, *H*				
c. H, *H*		*!		
d. L, *L*		*!		

[8] Paul de Lacy (p.c.) points out that one could specify the primary stem for *all* verbs. If the less marked stem happened to be specified as primary, as in a hypothetical /$\text{H}_{\text{primary}}$, LH/ listing, general markedness constraints ranked above FAITH-PRIMARY would stop it surfacing unchanged. The marking would thus be irrelevant in all cases except those where both allomorphs are equally marked.

Now consider a verb with only a singleton H tone (or a singleton L or ʔ), shown in (28). If anything other than H is used in either context, FAITH-IO will be violated, so it surfaces as H in all contexts.

(28) Single /H/ or /L/ or /ʔ/ will surface in both stems

{H}	*SECCONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
☞ a. H, *H*				
b. H, ʔ			*!	

On the other hand if the only listed allomorph is LH, shown in (29), it still cannot surface in secondary position because of undominated *SECCONTOUR. Since this dominates FAITH-IO, it will force changes in the input in secondary position, giving LH ~ H (or L). Such a stem would thus be indistinguishable from a verb with a /LH, H(or L)/ lexical entry.

(29) Single /LH/ cannot surface on secondary stem. No invariant LH stems.

{LH}	*SECCONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
☞ a. LH, *H*			*	*
b. LH, *LH*	*!			**

What about a putative lexical entry with a two-step difference, like /LH, ʔ/, shown in (30)? Such stems never surface, so MAX_{OO}-L&MAX_{OO}-H must be undominated. This constraint will look only within the output paradigm, and rule out the LH ~ ʔ output in (d). *SECCONTOUR will rule out using the LH input allomorph unchanged in secondary position, as in (c). Both (a) and (b) violate MAX-PARADIGM, so the decision is passed on to the general markedness constraint *CONTOUR, and our putative lexical entry would surface identically to an underlying /H, ʔ/ one. By Lexicon Optimization, then, the Zahao speaker would never postulate a /LH, ʔ/ entry in the first place. [9]

(30) Input paradigm with two-tone difference cannot surface; chainshift effect:

{LH, ʔ}	MAX_{OO}-L &MAX_{OO}-H	*SEC CONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
☞ a. H, ʔ			*	*	
b. LH, *H*			*	*	*!
c. LH, *LH*		*!			**

[9] A reviewer suggests that [ʔ, ʔ] is also a relevant candidate in (29). This is true, and it would be optimal. Such verbs exist, but the child would (by Lexicon Optimization) learn such a verb as /ʔ/, an input that is discussed in (28). More generally, winning candidates that are possible verbs of Zahao are not a problem for the analysis, even if they could have multiple sources. What matters is that the analysis never produces an impossible verb.

{LH, ʔ}	MAX_{OO}-L &MAX_{OO}-H	*SEC CONTOUR	MAX-PARADIGM	FAITH-IO	*CONTOUR
d. LH, ʔ	*!				*

Finally, consider an input with /H, ʔ/ (or /L, ʔ/), shown in (31). One last constraint must be added into our tableau to deal with this case. Such stems always surface with the glottal stop in the secondary stem, and the tone in the primary stem, suggesting that even simplex tones are avoided in secondary position. The necessary constraint *SECTONE must be ranked below FAITH-IO, since many secondaries do have surface H or L in secondary position, but nonetheless it plays a crucial role here in distinguishing between candidates (a) and (d), which differ only in the position in which the input H is realized.

(31) Positional markedness of single tones is lower-ranked, but stops tone addition in secondary stems:

{H, ʔ}	MAX-PARADIGM	FAITH-IO	*SECTONE	*CONTOUR
☞ a. H, ʔ				
b. H, *H*	*!		*	
c. ʔ, ʔ	*!			
d. ʔ, *H*			*!	

7. Conclusions and discussion of sections 1-6:

I have argued that the selection of allomorphs is controlled by four factors: (1) general phonological markedness, (2) positional markedness, (3) the pressure to realize all allomorphs in some context, and (4) the pressure to restrict intra-paradigm differences. If correct, this account has two consequences.

First, this analysis entails that H and L tones may tie on markedness, so that *H and *L are freely ranked with respect to each other, and there is no universally fixed ranking of *H >> *L, or vice-versa. This is a good result, since it allows for languages such as Zahao in which *H and *L are equally ranked; ones like the Bantu language Chichewa in which *L >> *H, and that thus have a H vs Ø contrast, and also ones such as Ruwund in which *H >>*L, and which have a L vs. Ø contrast. For discussion see Hyman (2001), Yip (2002).

Second, even heavily lexical alternations are subject to phonological and grammatical regulation. This observation is not new, but its documentation in a tonal language is of some interest. In the segmental domain, consider Spanish and English. Harris (1977) discusses alternating diphthongs in Spanish. The data in (32a) show alternating vowel/diphthong pairs. The data in (32b-c) show non-alternating vowels and diphthongs respectively:

(32)

Less - stressed	*Main stressed*	
a. n[e]gámos	n[yé]gan	'we-they deny'
p[o]démos	p[wé]den	'we-they can'
b. p[e]gámos	p[é]gan	'we-they arrive'
p[o]dámos	p[ó]dan	'we-they trim'
c. [ye]gámos	[yé]gan	'we-they arrive'
enc[we]rámos	enc[wé]ran	'we-they strip'

Harris says (1977:100): 'If one of these segments alternates, it is predictable what it alternates with..... Also, given an alternation, the diphthongs predictably appear under stress and the simple vowels unstressed. Diphthongization is thus a rule-governed phenomenon. It cannot be predicted, however, whether the vowels and diphthongs in question alternate at all.'[10]

Working in a rule-based framework, Harris uses these data to argue for the use of a lexical diacritic [D] that conditions a rule of diphthongization that applies only to stressed vowels in stems (not morphemes) bearing this diacritic. Under the OT approach advocated in this paper, alternating verbs would have all stem allomorphs listed, and a constraint would outlaw the more marked diphthong in the unstressed environment, but MAX-PARADIGM would require it to surface somewhere, i.e. in the stressed environment. For a somewhat different OT account, see Inkelas Orgun and Zoll (1997).

A second example can be drawn from English Latinate verb stems that show a t ~ s alternation. Other verbs, in the right-hand column, have fixed [t] or [s].

(33)					
	permit	permissive	vs.	devote	votive
	admit	admissive		promise	promissory
	remit	remissive			

Clearly the choice of [s] inter-vocalically is a sort of lenition, with continuants being less marked in that environment (Kirchner 1998) , but it cannot be phonologically derived synchronically. In our approach, the alternating verbs would have both allomorphs listed, and a constraint barring inter-vocalic [t] (a sort of positional markedness) would select [s] inter-vocalically, while MAX-PARADIGM would ensure that the [t] surfaced in some other context, such as word-finally.

Returning to tonal cases, Chen (2000: 46ff) points out that historical tones have often merged in modern dialects, but they may remain different in their sandhi behavior. The relevant cases are ones where the sandhi rule is clearly a natural rule, and yet where only a lexical subset of the possible inputs actually undergo that rule. As Chen makes abundantly clear, the historical roots of these splits are well known, but the issue here is how this is encoded for the modern speaker.

The Min dialect Quanzhou has a 31 citation phrase-final tone, where 5 is highest and 1 is lowest pitch. Some 31 words become 55 in non-final sandhi contexts, others become 11. Notice however that both level off, presumably a reduction in markedness. Chen suggests that for these tones alone, the sandhi forms are underlying and the citation tones derived. Since some other tones neutralize in sandhi environments, so that for example citation 24 and 22 both become 11, we cannot assume that the sandhi forms are underlying for all tones. In any case, there are good reasons for assuming that citation tones are underlying. The sandhi inventory [11, 44, 55, 35] is smaller than the citation inventory [22, 44, 55, 24, 31], and it has fewer contour tones, making it overall less marked. Since the pre-pausal forms are used in monosyllabic citation forms, which are uncontroversially prosodic heads, we must equate citation tones with head positions, and it follows that the sandhi forms must be in non-head positions. We know cross-linguistically that head position is where the maximum number of underlying contrasts are allowed to surface (Positional Faithfulness). It seems then that the 31 must be taken as underlying, and the difference in sandhi behavior attributed either to a diacritic, or to the presence in the lexical entry of different allotones:

---

[10] Albright, Andrade, and Hayes (2002) argue that in fact Spanish speakers make local generalizations about the likelihood of diphthongization, based on statistical regularities in the lexicon. Unfortunately, the available data for Zahao is not nearly extensive enough to test this possibility.

/31 ~ 11/ vs. /31 ~ 55/. In non-head position the less marked level tone will be chosen, with the more marked one chosen elsewhere, just as I have suggested for Zahao.

The strongest prediction of this approach is that if there are two allotones, and if sandhi take place in non-head position, then the sandhi one will always be the less marked of the two.[11] The weaker prediction is that in such systems the sandhi tone will never be more marked than the citation form. Of course, for full Min tone circles this is not true. Even in Quanzhou, for example, citation 55 becomes sandhi 35, which is presumably more marked. Given the right understanding of the system, though, this is not a counter-example. Suppose that the overall sandhi system encodes the changes somehow in the grammar, and that paired lexical allotones are limited to the idiosyncratic cases like /31 ~ 11/ vs. /31 ~ 55/. Then the prediction about no increase in markedness applies only to those cases, and as far as I know this is correct. Chen gives Wenling data in which citation 31 becomes either 13 or 31 before 31, and either 15 or 51 after 1ʔ. All the alternants here are arguably much the same in markedness, being contour tones. He also gives Pingyao data, in which citation 13 tones either stay unchanged , or become 31, and again there is no obvious change in markedness. Of course, this leaves untouched the larger and persistent problem of how to characterize the full tone circle in the first place. For a recent attempt, see Moreton (2004).

I have focussed so far on tonal changes, but there are also segmental and syllable structure changes, and the full flavour of the phenomenon cannot be grasped without a look at these facts. In the next (and final) section I will give an overview of these alternations. A formal analysis of these facts, however, must await future research. The central point is that here too we find a tendency towards unmarkedness, more specifically a move away from longer syllables and towards shorter syllables. A variety of different strategies are used to achieve this, as I shall now show.

8. Preliminary summary of segmental and syllable structure changes

Some of the segmental and syllable structure changes are independent of the tonal facts, while others are closely tied to the tonal changes. As an example of an independent change, all vowel-final primary stems add a final [t] in the secondary stem. Other changes are linked. If a L primary tone is unchanged, so is the syllable structure. But if the tone becomes H or ʔ, long vowels shorten. On the other hand H primaries that become L secondaries do not shorten. Let us look at all this in more detail.

8.1 Zahao syllable structure

Zahao has a vowel length contrast (shown here by primary stems):

(34)	baŋ L	‘resemble, look like’	baaŋ L	‘be tired’
	thri LH	‘marry’	thii H	‘bleed’
	fop H	‘suck; smoke’	doop H	‘jump down’

There are also phonetic length differences. LH syllables are significantly longer than level toned syllables. Glottalized syllables are almost certainly shorter. According to Osburne, the whole rhyme is glottalized, and sonorants are described as having “glottal cut-off”. Long vowels always shorten in glottal rhymes, but diphthongs may survive. There are no LH tones on glottalized or stop-final syllables.

---

[11] Obviously, some languages change the *head* tone. Mandarin, for example, changes the first of two L tones to LH in head position in response to the double pressure of the OCP and avoidance of L in heads. See Yip (2002). Here positional markedness does not prohibit a markedness increase (though positional faithfulness would!).

8.2 Secondary stem changes.

The key idea I wish to explore is that secondary stems tend to shorten, and never lengthen. The shortening may be neutralizing and phonological, or phonetic, or both. Recall that the context for secondary stems may be related to lack of focus - just the sort of context in which shortening would be unsurprising. Furthermore, cross-linguistically unstressed syllables tend to be shorter than stressed syllables. See Nooteboom (1997) for a good summary and references. Grouping these apparently quite different types of shortening together is of course controversial, but recent work has shown that phonology pays attention to quite fine-grained phonetic detail, and Zahao may be another such case. See Pierrehumbert (2002) and Gafos (2002) for discussion.

The schema below shows possible segmental shortenings that affect secondary stems, and in general they involve a single change, but the diagonal arrow cases involve two changes, mostly forced by the impossibility of long vowels in glottalized rhymes. t stands for final stops, n for final nasals, and y for final glides or liquids.

(35)

taa → taat		taay	taan
↘ ↘		↘	↓
ta → tat → taʔ		tay → tayʔ	tan

Alongside this, LH → L or H effectively shortens. Now let us look at some actual data. Unfortunately, we have no instrumental work on Zahao at present, but Maddieson (2002) has worked on Hakha Lai, and where relevant I include his length measurements for that dialect, and some measurements from other sources. The reader should bear in mind that we cannot be sure that the Zahao facts are the same.

In the following four sections I deal in turn with phonological vowel shortening, various phonetic shortenings, cases of length stability, and finally the lack of any instances of lengthening.

8.2.1 Markedness reduction:

The first set of data shows phonological removal of long vowels and diphthongs:

(36) a. Vowel shortening in secondary stems:

loom L	lom H	'celebrate'	cool L	colʔ	'rest'

b. Coalescence of rising diphthongs /ia/, /ua/ in secondary stems

nuam L	nomH	'enjoy'	lian L	len H	'be rich'

Cross-linguistically, diphthongs and contrastively long vowels are marked. The contrast is usually limited to stressed syllables, and unstressed vowels are all short. For Hakha-Lai, some average durations of phonologically long and short vowels before sonorants or oral stops are given below, showing that long vowels are more than twice as long as short vowels in both contexts.

(37) Hakha-Lai vowel durations (Maddieson 2002)

	Before [+son]	Before [-son]
Short	119	89
Long	277	250

8.2.2 Phonetic shortening:

The second set of data shows changes that result in phonetic shortening. See below for discussion.

(38) a. Loss of rising tones:

har LH	har L	'be difficult'
khaay LH	khaay L	'pull up, fetch'
phuum LH	phuum L	'bury'

b. Addition of glottalization:

bor H	bor?	'side with, support'
hruk H	hru?	'wear'
ney L	ney?	'have'

c. Syllable closure:

hoo H	hoot H	'preach'
hua H	huat H	'hate'

d. More than one of the above:

thri LH	thrit H	'marry'
diir L	dir?	'pull'
thoo H	thot L	'breathe'

Taking each of these in turn, starting with rising tones, Zhang (2000, 2002a,b) documents in great detail the cross-linguistic tendency for rising toned syllables to be longer (see also Gordon 1998). For example, in the unrelated languages Gã and in Mitla Zapotec rising tones (but not falling tones) are associated with lengthened syllables. Osburne specifically notes that this is true in Zahao. Loss of rising tones is therefore apparently accompanied by phonetic shortening. I should note, though, that Maddieson's data on Hakka Lai show LH essentially the same duration as L.

On glottalization, cross-linguistically syllables with glottal closure are normally shorter, and often non-moraic and not tone-bearing units (TBU's). They may behave as light for stress purposes, as they do in the South American language Cayapa (Lindskoog and Brend 1975). See also Zec (1988) on Kwakwala; Yip (2001) on Chaoyang. Roengpitya (1997) on Hakha Lai, says vowels before ? are about 75% of the length of comparable non-glottal vowels. Melnik (1997a) is less clear, but she explicitly says that the modal portion of the vowel before glottal closure is only about 80% the length of a non-glottal vowel. This is the portion available as a TBU. It is reasonable to conclude, then, that glottalization produces effective vowel shortening.

On the addition of a stop to open syllables, it is well-known that in many languages vowels in closed syllables tend to be shorter than their equivalents in open syllables (van Santen 1992). Among Asian languages, this effect has been documented for Cantonese, where the vowel in *taat* is about half the length of the vowel in *taa* (Kao 1971, Cheung 1986) and in the Chin language Hakha-Lai (Melnik 1997a) where long vowels are much shorter in closed syllables (200msec vs. 370msec). Maddieson's measurements for Hakha-Lai are less dramatic, but still substantial: 250ms vs. 334ms. Broselow, Chen and Hoffman (1997) agree, and go further. They offer evidence for several languages including Hindi and Malayalam that phonological structure and phonetic duration are a good match, since the structures motivated by the stress facts in a particular language are confirmed by its vowel duration data. In Hindi, vowels in open and closed syllables are the same length, so that VV = VVC and both are longer than V = VC. This means that VV always has two moras, and the C always has its own mora. Closed VC syllables therefore act as heavy. In Malayalam, on the other hand, vowels in closed syllables are shorter, so that VV is longer than VVC , which is longer than V which is longer than VC. They conclude that the C shares a mora with the preceding vowel, thus shortening the vowel. It follows that closed VC syllables should not act as heavy, and this is correct in Malayalam. For Zahao, things are less clear, because the minimal word in Zahao is CVV or CVC, suggesting that final codas may be moraic. Nevertheless, if Zahao is like the other Chin languages, then the addition of a consonant produces phonetic shortening in the vowel. This then makes it resemble Malayalam, with a non-moraic coda. One possibility is to

pursue a proposal of Gordon's for Khalkha Mongolian (Gordon 2002), and surmise that the minimal word in Zahao requires a certain amount of total energy in the rhyme, and that CVC syllables supply this, even if the vowel portion is somewhat shorter. I leave this question open for further research.

8.2.3 Markedness stasis:
The third set of data show that some secondaries are no shorter than their primaries:

(39) a. Vowels stay long:

keew H	keew H	'bite'
daar H	daar L	'entertain'
doop H	doop L	'jump down'
coom H	coom L	'raise, feed'

b. Diphthongs persist:

hua H	huat H	'hate'
cia?	cia?	'soak in water'
puar LH	puar L	'swell, be swollen'
niam H	niam L	'be low'

c. Modal, non-glottalized syllables persist:

pool LH	pool L	'associate, mix together'
faay H	faay L	'be clean'

8.2.4 No cases of markedness increases:
The final observation is that there are no cases of lengthening in secondaries:

(40) a. No vowel lengthening
b. No diphthongization
c. No loss of glottalization
d. No loss of coda consonants
e. No addition of rising tones.

8.3 Interdependence of tonal and segmental changes:
To some extent, the segmental shape of a syllable has no effect on its tonal behavior. For example, L ~ H pairs may have any segmental make-up[12], as may fixed non-alternating L or H stems. LH is never left unchanged, no matter what its syllable type. Nonetheless, there is an interdependence. The most obvious is that when tone is invariant, so is everything else: vowel length, glottalization, diphthongs , suggesting that stems must be marked as invariant *in toto*. The only exception is that V-final stems still add -t (historically a suffix).

Some of the interactions have straightforward causes. For example, L, H ~ ʔ pairings are only found if stems do not end in nasals, but this is because glottalization and nasalization are mutually exclusive in Zahao. Others are more obscure: LH ~ L and H ~ L exist only for C-final stems, and LH ~ H exists only for V-final stems. This cannot easily be attributed to an output prohibition of some sort, since the non-occurring secondaries are well-formed in other contexts.

One of the most interesting interactions is that between tone change and phonological vowel shortening. Length stays the same when the whole stem is invariant:

---

[12] Osburne says L ~H only for sonorant-final stems, but she gives examples with final stops: caakL~caakH 'be eager'

(41)	keew H	keew H	‘bite’
	cia?	cia?	‘soak in water’

and also when LH becomes H or L, but here the change from rising to level produces phonetic shortening anyway, and apparently this is phonologically sufficient:

(42)	cooy LH	cooy L	‘lift, carry’
	puar LH	puar L	‘swell, be swollen’

Length shortens when glottalization is added, because Zahao never allows long vowels in glottalized rhymes:

(43)	diirL	dir?	‘pull’
	kaapH	ka?	‘shoot’

The really interesting case is simple level-tone alternations. Shortening and coalescence happen in L ~ H cases, but not in H ~ L cases:

(44)	L ~ H	lian L	len H	‘be rich’
		loom L	lom H	‘celebrate’
	H ~ L	buan H	buan L	‘wrestle’
		coom H	coomL	‘raise, feed’

What this suggests is that when H primaries have L secondaries, they are sufficiently shorter phonetically by virtue of the tone change alone and no other shortening is needed. However when L primaries have H secondaries, they would be *longer,* so vowel shortening is essential to over-ride this. For this proposal to be correct, the intrinsic length of H tones must be longer than L tones. However, the cross-linguistic data is inconsistent, as the following summary shows:

(45)	Thai:	L longer than H	(Maddieson 1978, citing Gandour, n.d.)
	Taiwanese:	H longer than L	(Zee and Hombert 1976)
	Cantonese:	M is longest, H = L.	(Kong 1987)

For this story to go through, Zahao must thus be like Taiwanese, with H intrinsically longer than L, but unfortunately no phonetic data for Zahao is available to me.[13]

8.4 Summary:

In this final section, I have shown that syllable types in secondary stems are skewed towards shortness in three respects: all syllables are closed, vowels are mostly short, and many syllables are glottalized. I have suggested that phonological and phonetic shortness both matter, a claim that raises interesting issues about the phonetics- phonology boundary (Steriade 1999, 2000). The

[13] Larry Hyman (p.c.) points out that the shortening is usually accompanied by glottalization, except on nasals. These are cases where the Hakha-Lai cognate stem 1 is LH and Stem 2 is LH?, so historically the shortening may be attributed directly to this glottalization. (Nasal-final stems, which do not glottalize, get HL in his Falam data.) Synchronically, though, in Osburne’s data, shortening without glottalization is found, as shown by the first two examples in (36a).

pressure for shortness is held in check, however: shortening is never achieved by segmental deletion.

Analytically, these data pose a challenge. First, they cannot be attributed to an over-arching pressure for changing *something* (as in the Anti-faithfulness analysis of Hyman and VanBik (2002a)) since this would predict complementary distribution of laryngeal and length changes. Second, if we take a similar tack to that taken for the tonal changes, and propose handling the length alternations via allomorph listing, how do we capture the connections with tonal changes?

9. Conclusions

Zahao and the other Chin dialects are belatedly receiving considerable attention from theoretical linguists. The limited data available on Zahao makes the conclusions of this paper necessarily rather tentative, and it is to be hoped that other researchers will advance our knowledge of the phonetics, origins, cross-dialectal typology, and current status of these intertwined tonal and length alternations. Nonetheless, I believe I have established that a phenomenon which at first glance seems so heavily lexicalized as to be uninteresting in fact reveals itself as governed by fundamental phonological principles of markedness.

References

Adam, G. 2002. *From variable to optimal grammar: evidence from language acquisition and language change*. Ph.D. Dissertation, Tel Aviv University. ROA# 567-1202

Albright, A., A. Andrade, and B. Hayes. 2002. Segmental environments of Spanish diphthongization. Ms, UCLA.

Alderete, J. 1998. *Morphologically governed accent in Optimality Theory*. PhD Dissertation, U. Mass, Amherst.

Alderete, J. 2001. Dominance effects as trans-derivational anti-faithfulness. *Phonology*. 18.1: 201-253

Anderson, S. and W. Browne. 1973. On keeping exchange rules in Czech. *Papers in Linguistics VI* 4: 445-482.

Anttila, A. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory. 20.1: 1-42.*

Benua, L. 1997. *Transderivational identity: phonological relations between words.* PhD dissertation, UMass Amherst.

Broselow, E. , Su-I Chen, and M. Huffman. 1997. Syllable weight: convergence of phonology and phonetics. *Phonology* 14.1: 47-82.

Chen, M. 2000. *Tone sandhi.* Cambridge University Press. Cambridge.

Cheung, K-H. 1986. *The phonology of present-day Cantonese.* PhD dissertation, University College London.

Chhangte, T. 1989. Complementation in Mizo (Lushai). *Linguistics of the Tibeto-Burman Area* 12.1: 133-

Gafos, A. I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20.269-337.

Gandour, J. (n.d.) On the interaction between tone and vowel length. Cited in Maddieson (1978).

Gordon, M. 1998. The process-specific nature of weight: the case of contour tone distribution. *WCCFL* 18: .

Gordon, M. 2002. A phonetically driven account of syllable weight. *Language*. 78.1: 51-80.

Harris, J. 1977. Spanish vowel alternations, diacritic features, and the structure of the lexicon. In J. Kegl, D. Nash, and A. Zaenen, eds. *Proceedings of the Northeastern Linguistics Society* VII: 99-114.

Harris, J. and E-A Urua. 2001. Lenition degrades information: consonant allophony in Ibibio.

Speech, hearing, and language: work in progress. 13: 72-105.

Hartmann-So, H. 1989. Morphophonemic changes in Dai Chin. *Linguistics of the Tibeto-Burman Area* 1.2: 51-65.

Horwood, G. 2001. Antifaithfulness and subtractive morphology. ROA 466-0901.

Hyman, L. 2001. Privative tone in Bantu. In Shigeki Kaji (ed.) *Cross-linguistic studies of tonal phenomena*: 237-257. Tokyo: Institute for the Study of Langiuages and Cultures.

Hyman, L. and K. VanBik. 2002a. Tone and Stem-2 formation in Hakha-Lai. *Linguistics of the Tibeto-Burman Area*. 25.1.

Hyman, L. and K. VanBik. 2002b. Tone and syllable structure in Hakha-Lai. *Special Session, Tibeto-Burman/Southeast Asian Linguistics, BLS* 28.

Inkelas, S. O. Orgun, and C. Zoll. 1997. The implications of lexical exceptions for the nature of grammar. In Roca, I. ed. *Derivations and constraints in phonology.* Clarendon: Oxford. 393-418.

Kao, D. 1971. *The structure of the syllable in Cantonese.* The Hague: Mouton.

Kirchner, R. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27.2:341-149

Kirchner, R. 1998. *An effort-based approach to consonant lenition.* Ph.D. dissertation, UCLA.

Kong, Q-M. 1987. Influence of tones upon vowel duration in Cantonese. In *Language and Speech*, 30.4: 387-400.

Leben, W. 1973. *Suprasegmental Phonology.* PhD Dissertation, M.I.T.

Lehmann, F.K. 1996. Relative clauses in Lai Chin, with special reference to verb stem alternation and the extension of control theory. *Linguistics of the Tibeto-Burman Area* 19.1: 43-58.

Lindskoog, J. and R. Brend .1975. Fonemica Cayapa. In *Estudios fonologicas de lenguas vernaculas del Ecuador.* Quito: Ministerio de Educaion Publica.

Maddieson, I. 1978. Universals of tone. In Greenberg, J., Ferguson, C., and Moravscik, E. eds. *Universals of Human Language, Vol 2. Phonology.*: 335-366

Maddieson, I. 2002. Timing and alignment, a case study of Lai. Paper presented at the 8th International Symposium on Chinese Languages and Linguistics, Academia Sinica, Taipei.

McCarthy, J. 2002. Optimal paradigms. ROA# 485.

Melnik, N. 1997a. Sound system of Lai. *Linguistics of the Tibeto-Burman Area*. 20.2: 9-19.

Melnik, N. 1997b. Verbal alternations in Lai. *Linguistics of the Tibeto-Burman Area*. 20.2: 162-172.

Moreton, E. 2004. Non-computable functions in Optimality Theory. In J. McCarthy (ed) *Optimality Theory in Phonology*. Blackwell. Oxford. 141-164.

Nooteboom, S. 1997. The prosody of speech: melody and rhythm. In W. J. Hardcastle and J. Laver, (eds.), *The Handbook of Phonetic Sciences.* Oxford: Basil Blackwell, pp 640-74.

Osburne, A. 1975. *A transformational analysis of tone in the verb system of Zahao (Laizo) Chin.* Ph.D. Dissertation, Cornell University.

Ostapirat, W. 1998. Tiddim Chin tones in historical perspective. *Linguistics of the Tibeto-Burman Area* 21.1: 235-48.

Perlmutter, D. 1998. Interfaces: Explanation of allomorphy and the architecture of grammars. In S. PaPointe, D. Brentari, and P. Farrell, eds. *Morphology and its relation to phonology and syntax.* Stanford, CA: CSLI.

Pierrehumbert, P., M. Beckman, and R. Ladd. 2001. Conceptual foundations of phonology as a laboratory science., in *Phonological Knowledge,* ed. by Burton-Roberts, N., P. Carr, and G. Docherty, 273-304. Oxford: Oxford University Press.

Pierrehumbert, P. 2002. Word-specific phonetics, in *LabPhon 7,* ed. by Gussenhoven, C. and N. Warner, 101-40. Berlin: Walter de Gruyter.

Prince, A. and P. Smolensky. 1993. *Optimality theory: constraint interaction in generative grammar*. To appear, MIT Press. TR-2, Rutgers University Cognitive Science Center.

Available at http://ruccs.rutgers.edu/publicationsreports.html.
Reiss, C, M. Hale and M. Kissock. 1997. Output-output correspondence in Optimality Theory. ROA# 202-0697. To appear in Proceedings of WCCFL XVI.
Roengpitya, R. 1997. Glottal stop and glottalization in Lai connected speech. *Linguistics of the Tibeto-Burman Area*. 20.2: 21-56..
Silverman, D. 1997. Laryngeal complexity in Otomanguean vowels. *Phonology 14.2:235-262*
Steriade, D. 1999. Phonetics in phonology: the case of laryngeal neutralization. In *Papers in phonology 3 (UCLA Working papers in linguistics 2)*, ed. M. Gordon. pp25-145.
Steriade, D., 2000. Paradigm uniformity and the phonetics-phonology boundary. In *Acquisition and the lexicon (Papers in Laboratory Phonology 5)* ed., J. Pierrehumbert and M. Broe. Cambridge University Press. Pp 313-34.
Tranel, B. 1998. Suppletion and OT: On the issue of the syntax/phonology interaction. In E. Curtis, J. Lyle, and G. Webster, eds., *Proceedings of WCCFL*: 415-29.
van Santen, J.P.H. 1992. Contextual effects on vowel duration. *Speech Communication* 11: 513-46.
Weber, D. and W. Thiesen 2000 A synopsis of Bora tone. Ms, S.I.L.
Yip, M. 2001. Dialect variation in nasalization: Alignment or duration? In Xu, De-Bao, ed. *Chinese phonology in generative grammar.* Academic Press, London. pp 163-192. ISBN 0 -12-767670-8
Yip, M. 2002. *Tone*. Cambridge University Press.
Zec, D. 1988. *Sonority constraints on prosodic structure.* Ph.D. dissertation, Stanford University.
Zee, E. and J-M. Hombert. 1976. Duration and intensity as correlates of F_0. Paper presented to the 92nd meeting of the Acoustical Society of America, San Diego.
Zhang, Jie 2000. Phonetic duration effects on contour tone distribution. M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim (eds.) *Proceedings of NELS 30, Vol 2.* 775-785.
Zhang, Jie. 2002a. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. London: Routledge.
Zhang, Jie. 2002b. Contour tone licensing and contour tone representation. Paper presented at the 8th International Symposium on Chinese Languages and Linguistics, Academia Sinica, Taipei
Zoll, C. 1997. Conflicting directionality. *Phonology* 14: 263-86.

moira@linguistics.ucl.ac.uk
Department of Phonetics and Linguistics
and
Centre for Human Communication
University College London

C:\moira\zahao\Taiwan publish rev2.wpd