

**MODELING VARIATION
IN TAIWAN SOUTHERN MIN SYLLABLE CONTRACTION***

Yingshing Li and James Myers

ABSTRACT

In this paper we attempt to model variation in Taiwan Southern Min syllable contraction using the Gradual Learning Algorithm (GLA; Boersma and Hayes 2001), an Optimality-Theoretic model with variable constraint ranking. To explore the effectiveness of GLA, we look at three data sets of increasing complexity: non-variable fully contracted forms as analyzed by Hsu (2003), variable outputs as noted by Hsu and confirmed by other native speakers, and phonetically variable outputs collected in a speech production experiment by Li (2005). The results reveal that GLA is capable of providing plausible constraint ranking hierarchies that capture both major generalizations and variability. Stochastic constraint evaluation thus seems to be a promising mechanism in the construction of grammars.

1. INTRODUCTION

Researchers who have studied syllable contraction in Taiwan Southern Min (e.g. Cheng 1985; Chung 1996, 1997; Tseng 1999; Hsiao 2002; Hsu 2003) agree that it is fundamentally a variable phenomenon in at least three ways. First, it is variable across items: some syllable sequences are often contracted while others are unaffected. Second, it is even variable within items, which may appear in different forms on different occasions. Third, it is phonetically variable: sometimes syllable contraction is full, with two syllables being converted into one, and sometimes it is only partial, with deletion of a segment or two or lenition (e.g. shortening them or removing aspiration of the intervening consonants) rather than production of a single syllable. While much is

* This paper is an extended version of a chapter in Li (2005). We are grateful to two anonymous reviewers for helpful comments, although we are solely responsible for any inadequacies in the final product.

understood about the factors affecting syllable contraction, it would seem to be difficult to construct a model that can describe both systematic aspects and variability. Such a model may seem particularly difficult to construct given one of the fundamental goals of generative linguistics since Chomsky (1965), namely to explain how it is that children manage to acquire linguistic systems.

In this paper we test an Optimality-Theoretic model that was developed to handle problems of exactly this kind: the Gradual Learning Algorithm (GLA; Boersma and Hayes 2001; see also Boersma 1998; Apoussidou and Boersma 2004). Similar to the OT learning algorithm of Tesar and Smolensky (2000), GLA is a fully automatic model of child language acquisition, taking language data to generate a hypothesized grammar; however, unlike Tesar and Smolensky's model, GLA is able to handle variable data. Thus if the adult language shows a variable pattern, where variant A appears 30% of the time while variant B appears 70% of the time, GLA will be able to acquire this pattern from raw data. As sociolinguists have known for a long time (see review in Labov 1994), this is something that children seem able to do, learning not only which phonological forms to change into what in which context, but also learning what proportion of the time they should do it. Moreover, GLA is potentially capable of describing all three types of variability, although in this paper we will primarily be concerned with the second type (multiple output forms for any given input form).

The modeling procedure we carry out in this paper comprises three major steps. First, we determine the OT constraints for syllable contraction. Here we build on insights from previous derivational models, particularly the Sonority model of Hsu (2003), as well as formalisms developed in the OT framework by Hsiao (2002) and Hsu (2005). Second, we prepare the learning data for modeling. To explore the strengths and weaknesses of the GLA model, we look at three data sets of increasing complexity: the categorical fully contracted forms that are the focus of Hsu's (2003) analysis, then the variable outputs as noted by Hsu (2003) as confirmed by native speakers whom we consulted, and finally the phonetically variable outputs collected in a speech production experiment by Li (2005). Third, we input the learning data into the Gradual Learning Algorithm in order to construct the appropriate grammar. Since all three data sets represent the same language, a primary concern is whether GLA learns the same grammar from them all.

This paper is organized as follows. In Section 2, we introduce the principles governing Taiwan Southern Min syllable contraction. In

Section 3, we formalize these governing principles as OT constraints. In Section 4, we describe the procedures used to carry out the GLA modeling and then evaluate the results. In Section 5, we summarize our conclusions.

2. PRINCIPLES OF TAIWAN SOUTHERN MIN SYLLABLE CONTRACTION

Taiwan Southern Min syllable contraction has long attracted the attention of phonologists, starting particularly with Cheng (1985). Aside from a few cases where its effects have been fully lexicalized (e.g. *gua* + *-n* > *guan* ‘we’; see Tseng 1999 for discussion of its diachronic effects), syllable contraction is always optional, with its probability of application depending on factors such as segmental makeup, rhythmic pattern, prosodic boundary, lexical frequency, lexical category, and speaking rate (Tseng 1999). Moreover, when it does occur, syllable contraction can apply in more than one way, generating more than one possible output for any given input. Examples showing this variability are given in (1) (all from Hsu 2003 except for the second alternatives in (1e-f), which were confirmed with native speakers). Note that since our focus in this paper is the segmental changes in syllable contraction, we do not transcribe tone, which of course also contracts.

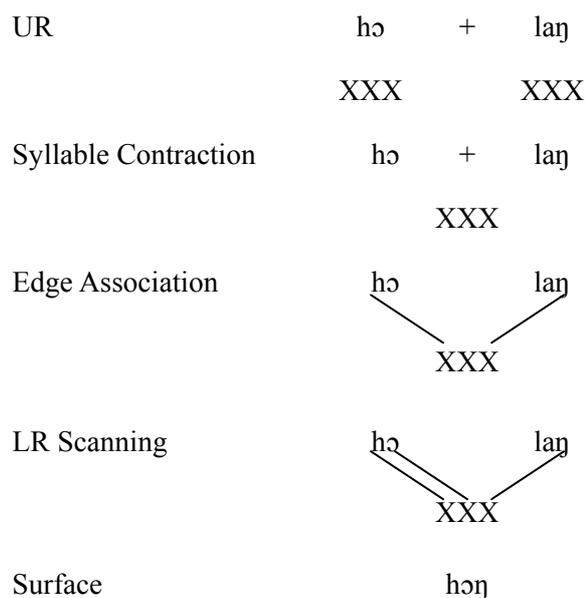
(1) Examples of Taiwan Southern Min syllable contraction

a. <i>bo</i> + <i>e</i>	<i>bue</i> / <i>be</i>	‘unable’
b. <i>si</i> + <i>tsun</i>	<i>sin</i> / <i>sun</i>	‘moment’
c. <u><i>tsa</i></u> + <u><i>bɔ</i></u> <i>laŋ</i>	<u><i>tsau</i></u> / <u><i>tsɔ</i></u> <i>laŋ</i>	‘woman’
d. <i>k^hi</i> + <i>lai</i>	<i>k^hiai</i> / <i>k^hai</i>	‘get up’
e. <i>hɔ</i> + <i>laŋ</i>	<i>hɔŋ</i> / <i>haŋ</i>	‘by someone’
f. <u><i>bɔ</i></u> + <u><i>iau</i></u> <i>kin</i>	<u><i>bua</i></u> / <u><i>bau</i></u> <i>kin</i>	‘it doesn’t matter’

The formal generative analysis of Taiwan Southern Min syllable contraction began with the derivational Edge-In model of Chung (1996, 1997). Adopting the notion of Edge Association (Yip 1988), this model proposes three key principles governing syllable contraction: each syllable has a prosodic template of three X-slots on the skeletal tier, the association between the melodies and the X-slots proceeds with both edges of the template (Edge Association), and the association begins from left to right for the medial X-slot (LR scanning). Chung also noted

that the surface form is usually required to be a grammatical syllable, that is, one that obeys all of the other phonotactic constraints of the phonological system. Figure (2) illustrates Chung's model. In the underlying representation, two syllables *hɔ* and *laŋ* belong to separate prosodic templates. After the contraction process, two prosodic templates merge into one. Edge Association requires the marginal melodies to associate with the prosodic template. LR scanning selects the leftmost vowel *ɔ* as the nucleus of the contracted syllable instead of the vowel *a*. As a result, the surface form is *hɔŋ*. The logically possible alternative output **huaŋ* (allowed by Chung's Vowel neutralization, in which *ɔ* can be transformed to *u*) is ruled out because it violates Taiwan Southern Min phonotactics (specifically, Chung's Branching-N constraint which bans the co-occurrence of a prevocalic *u* and a dorsal coda).

(2) *hɔ + laŋ > hɔŋ* 'by someone'



Chung required one proviso for the LR scanning: if the second syllable ends with a high vowel, it links to the marginal X-slot assuming that it is underspecified for [+consonantal]. For example, the high vowel *i* in *tsa + k^hi > tsai* 'morning' links to the rightmost X-slot in the same way as a coda consonant in the Edge association. However, if the second

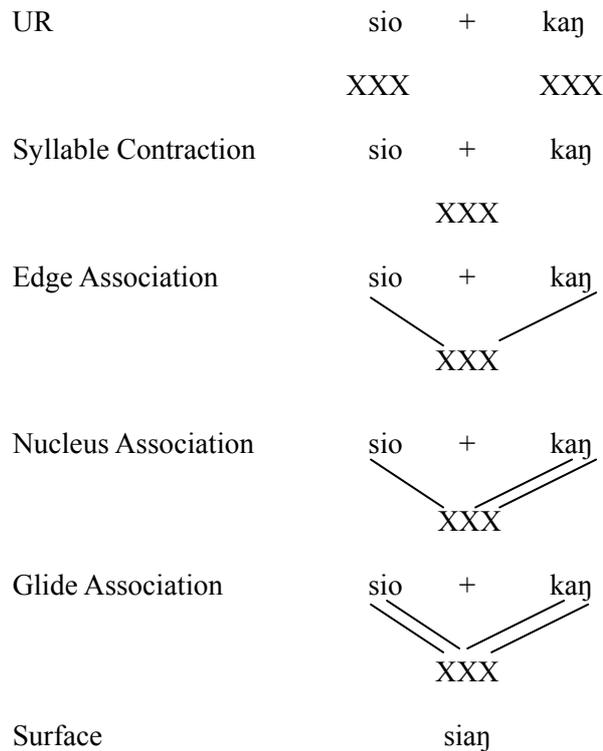
syllable ends with a [+syllabic] segment (i.e. a non-high vowel), it receives priority to associate with the medial X-slot among the other vowels such as *bo + e > be* ‘unable’.

This additional element of complexity suggests that the process of contracting nuclei is crucially sensitive to sonority, an insight on which Hsu (2003) elaborated in the construction of the Sonority model. Hsu modified the Edge-In principle by proposing that it affected the first syllable onset and the second syllable consonant coda alone, so that only consonants can occupy the marginal X-slots. With the assumption that all vowels are linked to the medial X-slot, this model is thus able to provide a single account for *tsa + k^hi > tsai* and *bo + e > be* (or *bue*, as in Hsu’s dialect). Building on the XXX syllable model of Chung, Hsu added three new syllabic principles. First, the construction of the contracted syllable begins with the linking of the N (nucleus) to the central X slot, followed by the formation of rising diphthongs, and finally the formation of falling diphthongs. Thus, the model seems to favor rising over falling diphthongs, a point to which we return later. Second, selection of the vocoid is determined by the (partially language-specific) sonority hierarchy (a > ɔ > e > o > i > u). Note that this hierarchy may help explain why for some of the speakers whom we consulted, the preferred contracted form for *hɔ + laŋ* is *haŋ*, not *hɔŋ* as predicted by the LR scanning principle of Chung’s model in (2): /a/ is more sonorous than /ɔ/. A certain amount of the LR scanning principle remains in Hsu’s model, since if there is a tie in sonority between possible choices from each of the two source syllables, the leftmost one is favored, but this claim is a difficult one to test since vocoids will only have identical sonority if the vocoids themselves are identical (as we will see below, the centrality of sonority affects the interpretation of LR scanning in the OT model as well). Third and finally, the syllable constructed by syllable contraction must obey the Maximality Principle, as long as it also observes phonotactic constraints.

Figure (3) provides an illustration of Hsu’s model. First, two prosodic XXX templates are combined into one. In Edge Association, the marginal consonants are associated with the first and last X-slots. In Nucleus Association, the most sonorous vowel, *a*, takes the priority in docking at the medial X-slot. Finally, the vowel *i* associates with the medial X-slot to construct the maximal diphthong *siaŋ*, observing phonotactic constraints (by contrast, Chung’s LR scanning would wrongly predict *sioŋ*). The logically possible alternative output **suaŋ* (after applying vowel neutralization) is ruled out because it violates

phonotactics (this time the Branching-R constraint, which blocks co-occurrence of a high vowel *u* and a dorsal consonant *ŋ* in the VC-structured rime).

(3) *sio* + *kaŋ* > *siaŋ* ‘the same’

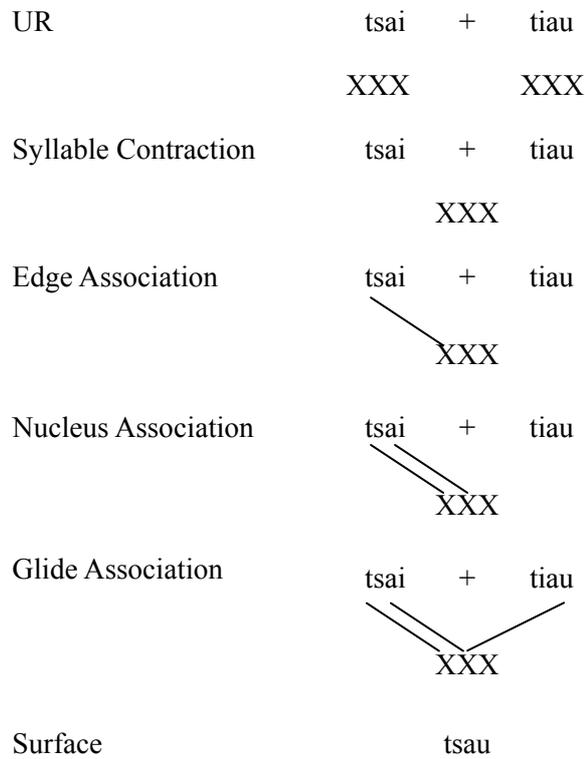


Hsu (2003) also proposed three additional filters on outputs. The first, the No Crossing Line Constraint, bans reversing the order of association between the melodic and skeletal tiers. The second filter, the Non-Identity Constraint, prohibits the total identity (both segmental and tonal) between the contracted syllable and either of the source syllables. Figure (4) illustrates the operation of these two constraints. In Nucleus Association, the leftmost vowel *a* docks at the medial X-slot, and then in Glide Association, the vowel *u* is also linked, thereby constructing the maximal diphthong *au*, and as consistent with phonotactic constraints. The alternative output **tsiau* is ruled out because after the placement of

Modeling Variation in Syllable Contraction

the leftmost vowel *a*, the vowel *i* on the right side would have to cross the association line of the vowel *a*, thus violating the No Crossing Line Constraint. The alternative output **tsai* is ruled out because it is identical with the first source syllable, violating the Non-Identity Constraint.

(4) *u tsai21 + tiau21 k^hi > u tsau21 / *tsiau21 / *tsai21 k^hi* ‘be able to go’

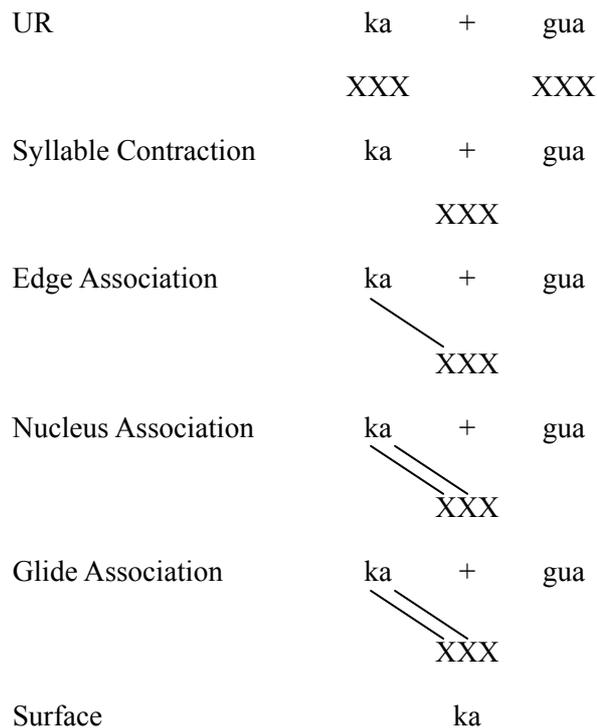


The third filter on outputs proposed by Hsu is Glide Transfer, which requires input-output structural correspondence on the part of glides. That is, a prevocalic glide must be a prevocalic glide after the contraction process, while a postvocalic glide must be postvocalic one. Figure (5) illustrates how Glide Transfer and the No Crossing Line Constraint block out the alternative licit syllables. The unattested **kau* violates Glide Transfer because the prevocalic glide *u* of the source syllable becomes postvocalic in the output, while **kua* violates the No Crossing Line Constraint because of the reversed association between

Yingshing Li; James Myers

the melodic and skeletal tiers.

(5) ka21 + gua33 ma > ka23 / *kau23 / *kua23 ma ‘scold me’



In this section, we have introduced Chung’s (1996, 1997) and Hsu’s (2003) models governing Taiwan Southern Min syllable contraction. Notice that neither model directly addresses the problem of variability, though in principle both could: presumably a given input can generate a variety of outputs if all of them obey the principles and constraints affecting the derivations. However, it seems worthwhile to study whether the key insights of these models can be formalized in an OT approach. This is not only necessary if we are to use GLA to solve the variability problem, but it seems likely that OT can provide an appropriate formalism for dealing with certain phenomena that are otherwise mysterious from a derivational perspective. In particular, Hsu refers to several constraints, some of which seems to be stronger than others; this suggests an analysis involving constraints that are ranked and violable.

3. OT CONSTRAINTS FOR TAIWAN SOUTHERN MIN SYLLABLE CONTRACTION

There has been no previous attempt to model Taiwan Southern Min syllable contraction within an OT formalism with anywhere near the degree of detail of the derivational analysis of Hsu (2003). Thus before demonstrating the application of the GLA learning algorithm, we first describe a more “traditional” OT analysis. This analysis builds primarily on the insights of Hsu (2003), but some elements match those proposed in the OT formalism of Hsu (2005) (primarily addressing syllable contraction in Cantonese, but with a brief mention of Taiwan Southern Min) and Hsiao (2002) (addressing the tone contraction that accompanies syllable contraction in several Sinitic languages, including Taiwan Southern Min).

Note first that fully contracted syllables always fit into the maximal syllable template allowed in Taiwan Southern Min, and they always preserve the marginal consonants. For simplicity of exposition, we assume the existence of undominated constraints that guarantee these generalizations, and so consider only output candidates that obey them.

Our analysis proper begins at the heart of Taiwan Southern Min syllable contraction: sonority. According to the Sonority Sequencing Principle (Venneman 1972, Selkirk 1984, Clements 1990), syllable margins (onsets and codas) prefer segments of low sonority while syllable nuclei prefer segments of high sonority, following the universal sonority hierarchy vowels > glides > liquids > nasals > fricatives > stops. Considering only the vocoids in Taiwan Southern Min, Hsu (2003) proposed, as we saw above, the sonority hierarchy $a > \text{ɔ} > e > o > i > u$, which we convert into a family of ranked constraints, as described below. However, our analysis is also designed to be flexible enough to accommodate partial contraction, where an intervening consonant in a disyllabic sequence may remain, and thus we require constraints that refer to the logical possibility of consonantal nuclei as well (which do occur cross-linguistically, as in the Imdlawn Tashlhiyt dialect of Berber analyzed in OT terms in Prince and Smolensky 2004). Here we only refer to two categories of consonants: C_1 indicates the coda of the first syllable and C_2 the onset of the following syllable. The relative appropriateness of nuclei (in terms of their sonority) is thus expressed by the constraint family in (6). Note that the ranking indicates that /a/ is the

best possible nucleus.

- (6) a. *NUC/ α The segment α cannot be in nucleus position.
 b. { *NUC/C₁, *NUC/C₂ } » *NUC/u » *NUC/i » *NUC/o » *NUC/e » *NUC/ɔ » *NUC/a

In (7), the vowel *a* is more sonorous than the vowel *o*, thus producing the output *siaŋ* instead of *sioŋ*. Here we assume, following Chung (1996) and Hsu (2003), that glides are linked to the nucleus slot along with the vowel.

(7) *sio* + *kaŋ* > *siaŋ* ‘the same’

sio + kaŋ	*NUC/i	*NUC/o	*NUC/a
sioŋ	*	*!	
☞ siaŋ	*		*

Importantly, for many items more than one output is possible. For example, *si* + *tsun* ‘moment’ is usually contracted as *sin*, but it sometimes also appears as *sun*. The first output is expected, since /i/ is assumed to be more sonorous than /u/, but the second output is not. This raises the possibility that the sonority hierarchy in Taiwan Southern Min is not perfectly fixed, but instead is allowed to vary, at least somewhat. Note that this claim is in principle compatible with the notion of language-specific constraint rankings; if children acquiring different languages can learn different hierarchies, why can they not also learn about variation in hierarchies within a language?

This kind of variability is simple to express in OT, namely as variable ranking, a notion first formalized in print in Anttila (1997) and greatly expanded on in Boersma (1998). The technical details as to how this works in Boersma’s model will be discussed later; for the present, we simply present it in (8) showing alternative rankings and their outputs.

(8) *si + tsun* > *sin / sun* ‘moment’

a.

<i>si + tsun</i>	*NUC/u	*NUC/i
☞ <i>sin</i>		*
<i>sun</i>	*!	

b.

<i>si + tsun</i>	*NUC/i	*NUC/u
<i>sin</i>	*!	
☞ <i>sun</i>		*

Note that since /u/ appears in the second syllable in the input, we cannot assume that the alternative form *sun* appears because of a preference for vowels in the first syllable, as would be implied by the LR scanning principle of Chung (1996, 1997). Nevertheless, in other cases of variability it seems that just such a principle is necessary. To formalize left-to-right linking in a nonderivational OT approach, we adopt the faith constraint ANCHOR(L,V), which requires that the output preserve the leftmost vowel of the input (i.e. the vowel of the first syllable); see McCarthy and Prince (1995) for a description of the first use of such a constraint.

(9) ANCHOR(L,V)

The leftmost vowel of the input (syllable sequence) must have a correspondent in the output (contracted form).

The application of this constraint is shown by variable patterns such as *hɔ + laŋ* > *hɔŋ / haŋ* ‘by someone’. As noted earlier, in most cases the output is *hɔŋ*, which violates the sonority hierarchy $a > \text{ɔ}$, but the output *haŋ* is not impossible, at least for some speakers.¹ While we could assume that this variability derives from the variable ranking of *NUC/ɔ and *NUC/a, a positing of the variable ranking of ANCHOR(L,V) allows us to capture the generalization that, in most cases, it is indeed the

¹ An anonymous reviewer suggests that in most exceptions to Hsu’s (2003) sonority model, the output onset comes from the first syllable and the rime from the second syllable, as in sonority-violating forms such as *sun* from *si + tsun* ‘moment’. Assuming that this observation is valid, we would need to add a faith constraint referring to prosodic units, something such as IDENT(rime), though this may conflict with the standard OT assumption that the input contains no prosodic structure (see also discussion of LINEARITY below).

leftmost vowel that is preserved. This generalization is particularly striking in the case of $h\text{ɔ} + laŋ$, which normally generates an output that violates the sonority hierarchy. If we assume that the common output form $h\text{ɔ}ŋ$ is harmonic (i.e. the unmarked form preferred by the OT grammar), we require a constraint that can force this output, and that constraint is ANCHOR(L,V). Illustrative tableaux are shown in (10).

(10) $h\text{ɔ} + laŋ > h\text{ɔ}ŋ / haŋ$ ‘by someone’

a.

$h\text{ɔ} + laŋ$	*NUC/ɔ	*NUC/a	ANCHOR(L,V)
☞ $haŋ$		*!	
$h\text{ɔ}ŋ$	*!		*

b.

$h\text{ɔ} + laŋ$	ANCHOR(L,V)	*NUC/ɔ	*NUC/a
$haŋ$	*!		*
☞ $h\text{ɔ}ŋ$		*	

Somewhat more conclusive evidence for the function of constraint ANCHOR(L,V) is shown by variability in $tsa + b\text{ɔ} laŋ > tsau / ts\text{ɔ} laŋ$ ‘woman’. On the assumption that the /u/ of $tsau$ is “actually” /ɔ/ (in some sense), we cannot say that the variation here occurs due to variable ranking of *NUC/ɔ and *NUC/a, since *NUC/ɔ is violated by both outputs ($tsau$ and $ts\text{ɔ}$), making the ranking of these two constraints irrelevant. This means that $tsau$ must surface when it does by virtue of the occasionally higher ranking of ANCHOR(L,V), as shown in (11).

(11) $tsa + b\text{ɔ} laŋ > tsau / ts\text{ɔ} laŋ$ ‘woman’

a.

$tsa + b\text{ɔ}$	ANCHOR(L,V)	*NUC/ɔ	*NUC/a
☞ $tsau$ ($tsa\text{ɔ}$)		*	*
$ts\text{ɔ}$	*!	*	

b.

$tsa + b\text{ɔ}$	*NUC/ɔ	*NUC/a	ANCHOR(L,V)
$tsau$ ($tsa\text{ɔ}$)	*	*!	
☞ $ts\text{ɔ}$	*		*

Of course, there is a problem when we try to flesh out the assumption that the glide in $tsau$ is /ɔ/ “in some sense,” since OT does

not have the luxury of invoking multiple derivational levels. If the output form really contains a phonetic [u], then the ranking of *NUC/ɔ and *NUC/a is not relevant after all, so we can not use this kind of example to argue for the necessity of ANCHOR(L,V). Fortunately we can otherwise ignore vowel neutralization (i.e. the transformation of a mid vowel to a high vowel during diphthongization) in our OT analysis, since a glide always has a lower sonority than the adjacent nucleus in a diphthong.

Regarding diphthongization, recall that Hsu (2003) proposed that the construction of rising diphthongs preceded that of falling diphthongs in the realization of a contracted syllable. We formalize this phenomenon in terms of the constraints *FALLING and *RISING, referring to sonority (e.g. /ia/ has rising sonority while /ai/ has falling sonority).

- (12) *FALLING
Falling diphthongs are disallowed.
- (13) *RISING
Rising diphthongs are disallowed.

Hsu's claim of a preference for rising diphthongs implies the ranking *FALLING » *RISING. Since diphthongization is neutral with respect to ANCHOR(L,V) and the constraint family *NUC/V, the ranking at this point is as follows: {[*FALLING » *RISING], ANCHOR(L,V), *NUC/V}. Note that the ranking *FALLING » *RISING seems to be independently motivated by the same forces that give rise to the constraints ONSET and NOCODA. That is, ONSET indicates a preference for [CV] over [V], meaning a preference for rising sonority over level sonority, while NOCODA indicates a preference for [V] over [VC], meaning a preference for level sonority over falling sonority. These familiar constraints and our proposed ranking *FALLING » *RISING thus conspire to produce syllables that start with a bang but end with a whimper, so to speak. The only problem, as we will see, is that GLA does not induce this ranking from our data. There seem to be two reasons for this, both implicit in our discussion so far.

To see this, consider the examples in (14).

- (14) a. *bo* + *iau kin* > *buə* / *baʊ kin* (cf. **bia kin*) 'it doesn't matter'
b. *ke* + *lai* > *kai* / *kiai* (cf. **kia*) 'come over'
c. *k^hi* + *lai* > *k^hai* / *k^hiai* (cf. **k^hia*) 'get up'

According to the derivational account of Hsu (2003), *bua* in (14a) is derived as follows. First the most sonorous vowel *a* docks at the medial X-slot, and then the prevocalic glides (*o* and *i*) compete to construct a rising diphthong. The sonority harmonic hierarchy $o > i$ determines the winner *oa* (neutralized as *ua*) via Glide Association. The difficulty for this account is that native speakers also occasionally pronounce a falling diphthong *bau* (a few of the native speakers whom we consulted reported that *biau* and *bua*u** are also possible outputs, a further complexity we set aside). Tableau (15) shows how such variation can be handled through variable ranking of the two diphthong constraints.

(15) *bo* + *iau* *kin* > *bua* / *bau* ‘it doesn’t matter’

a.

bo + iau	*FALLING	*RISING
☞ <i>bua</i>		*
<i>bau</i>	*!	

b.

bo + iau	*RISING	*FALLING
<i>bua</i>	*!	
☞ <i>bau</i>		*

In examples like those in (14b-c), a purely rising diphthong is never created (the triphthong /iai/ is also possible). This may suggest that for some items the two diphthong constraints have their ranking fixed in the opposite way (i.e. *RISING » *FALLING). Another possibility (which we do not pursue here) is that there is another ANCHOR constraint requiring a string-final /i/ to remain in the output. Moreover, the same ANCHOR(L,V) constraint that we had trouble motivating above duplicates some of the work of *FALLING. This is clear from examples such as *bo* + *iau* *kin* > *bua* *kin* ‘it doesn’t matter’, where the output syllable *bua* obeys both constraints (assuming identity between /o/ and /u/ as above). The consequences for the operation of GLA will be discussed below.

Since variable syllable contraction involves variable deletion, we require faith constraints to block deletion; we simply adopt MAX-IO(C) in (16) and MAX-IO(V) in (17).

(16) MAX-IO(C)

Consonants in the input must have correspondents in the output.

(17) MAX-IO(V)

Vowels in the input must have correspondents in the output.

Since full contraction generally involves the deletion of consonants but not necessarily the deletion of vowels, the ranking MAX-IO(V) » MAX-IO(C) seems plausible. Of course, this ranking can only be learned if there are possible output candidates that retain intervening consonants, but such candidates are at best partially contracted. Thus for the data examined by Hsu (2003), the ranking here is essentially irrelevant. However, we will assume that MAX-IO(V) is outranked by ANCHOR(L,V) in order to prevent the prevocalic glide of the second syllable from being preserved in cases such as *u tsai + tiau k^hi > u tsau k^hi* ‘be able to go’ (cf. **u tsiau k^hi*). Another possibility would be to generalize the constraint we propose below for handling Glide Transfer, but we do not pursue it here. Notwithstanding the above points, as with all of our “hand-rankings”, the practical test will be to see what GLA induces from the data.

We now turn to the additional filters proposed by Hsu (2003). Following the notation of Hsu (2005) in her OT analysis of Cantonese syllable contraction, we incorporate all phonotactic constraints into a single cover constraint PHONOTACT.

(18) PHONOTACT

The output must observe phonotactic constraints.

This constraint stands for a wide variety of constraints encoding at least the observations of Chung (1996). Thus as described above, the Branching-R Constraint bans [+high][+high] in the VC-structured rime, and the Branching-N Constraint prevents a prevocalic *u* from co-occurring with a dorsal coda. The Dissimilatory Constraint bans [αback](...)[αback] within the nucleus. The N-Constraint requires diphthongs to have at least one high vowel. The Coda Condition demands that codas must be oral/nasal stops. The Labial Constraint prohibits [+labial](...)[+labial] within the syllable unless the two labials are onset and nucleus. The One-nasal Constraint stipulates that a maximum of one nasal autosegment may occur in a syllable.

In principle, the constraint PHONOTACT should be undominated because even the Maximality principle (constructing a maximal syllable) must observe phonotactics. This implies the ranking PHONOTACT » {[ANCHOR(L,V) » MAX-IO(V) » MAX-IO(C)], [*FALLING » *RISING],

*NUC/V}.

However, as many researchers have noted, contracted syllables do not always observe phonotactics in Taiwan Southern Min (Tseng 1999 points out that this property could be a diachronic source for new additions to the syllabary). Examples of contracted syllables that violate phonotactic constraints are shown in (19). Thus the output of (19a-b) contains the rime /oi/, otherwise unattested in the Taiwan Southern Min syllabary, and (19c-d) contains the sequence /iɔ/, similarly unattested. The triphthongs in (19e-f) are also disallowed in nonderived syllables.

- (19) a. *loʔ + k^hi > loi* ‘get down’
 b. *toʔ + ui > toi* ‘where’
 c. *he + ɔ > hiɔ* ‘interjection for a sudden realization’
 d. *si + bɔ > siɔ* ‘right?’
 e. *ke + lai > kiai* ‘come over’
 f. *k^hi + lai > k^hiai* ‘get up’

In dealing with these cases, it seems that we must demote the ranking of PHONOTACT below MAX-IO(V), as shown in (20) and (21).

- (20) *si + bɔ > siɔ* (cf. **si* / **sɔ*) ‘right?’

si + bɔ	ANCHOR(L, V)	MAX-IO (V)	PHONOTACT	*FALLING	*RISING	*NUC/i	*NUC/ɔ
siɔ			*		*	*	*
si		*!				*	
sɔ	*!	*					*

(21) $k^hi + lai > k^hai / k^hi ai$ ‘get up’

$k^hi + lai$	ANCHOR(L,V)	MAX-IO(V)	PHONOTACT	*FALLING	*RISING	*NUC/i	*NUC/a
$k^hi ai$			*			**	*
$k^h ai$	*!	*		*		*	*
$k^h ia$		*!			*	*	*
$k^h a$	*!	**					*
$k^h i$		*!*				*	

Finally, we consider the No Crossing Line Constraint, Glide Transfer, and the Non-Identity Constraint. We reinterpret No Crossing in terms of the anti-metathesis constraint LINEARITY, which bans reversing the order of segments, as in $tsa + k^hi > tsai$ (cf. $*tsia$) ‘morning’.

(22) LINEARITY

The linear order of segments in the input is maintained in the output.

LINEARITY differs from Hsu’s (2003) No crossing Line Constraint in that the former does not invoke autosegmental association lines. As a faith constraint referring to sequential position, it also interacts in some cases with ANCHOR(L,V). For example, in $u tsai + tiau k^hi > u tsau k^hi$ (cf. $*u tsiau k^hi$) ‘be able to go’, Hsu’s account rules out $tsiau$ because the leftmost /a/ is chosen in Nucleus Association to break the tie between the two identical /a/ vowels, making it impossible for the /i/ of the second syllable to be linked across the association line of /a/. By contrast, in our analysis LINEARITY cannot rule out $tsiau$ by itself since the output /a/ could come from the second syllable, thereby allowing the preservation of the linear sequence /iau/. However, this would represent a violation of ANCHOR(L,V), which requires the output /a/ to be the correspondent of the first /a/, not the second one.

Hsu’s (2003) Glide Transfer requires that a prevocalic (postvocalic) glide must remain a prevocalic (postvocalic) glide after the contraction process, as in $ka + gua ma > ka ma$ (cf. $*kau ma$) ‘scold me’; note that in the unattested $*kau ma$, the original /u/ is preserved only by changing it from a prevocalic to a postvocalic glide. This is somewhat difficult to

formalize in OT, since it seems to require the preservation of a position defined in terms of syllable structure, not merely sequential order. Yet like the underlying representation in derivational theories, the input in OT is generally assumed to contain no prosodic structure. A solution is to follow Hsiao (2002) in his analysis of tone contraction and adopt base-derivative (BD) correspondence. That is, we treat the contracted form as “morphologically” derived from the original uncontracted syllable sequence. In this case, we may then say that it is not the glide of the input that is maintained in the contracted form, but rather the glide in the surface form of the uncontracted form. We capture this with the constraint LINEARITY-BD(G,V), which preserves the sequential order of glides (however they may best be defined) and vowels between the base (surface uncontracted) and derived (surface contracted) forms.²

(23) LINEARITY-BD(G,V)

The relative position of the glide and vowel within a diphthong must be consistent between base and derived form.

Seeing syllable contraction as being similar to a morphological process also aids in the OT formalization of Hsu’s (2003) Non-Identity principle, which prohibits total identity between the contracted syllable and either the source syllables, as in *u tsai + tiau k^hi > u tsau k^hi* (cf. **u tsai k^hi*) ‘be able to go’. This constraint has an obvious benefit for the listener, in that it makes it possible to reconstruct the intended morphemes. A very similar notion has been formalized in the OT literature on the phonology-morphology interface in the form of anti-faithfulness constraints, first proposed by Alderete (2001), which require non-identity between base and its morphologically derived form; Hsiao (2002) also made a similar connection in his analysis of tone contraction. Here we simply stipulate a constraint NON-IDENTITY as in (24).

(24) NON-IDENTITY

The output must not be totally identical with either syllable of the base (inclusive of syllable structure and tone).

Hsu suggested that these last three constraints should be obeyed even

² By contrast, LINEARITY in (22) can still be assumed to involve input-output correspondence, since it does not refer to prosodic structure.

when this involves violation of phonotactic constraints, since these constraints are never violated while phonotactic constraints sometimes are (Hsu 2003:374). This suggests the constraint ranking {LINEARITY, NON-IDENTITY, LINEARITY-BD(G,V)} » PHONOTACT, though, as Hsu warns, taking this as a fixed ranking may cause ranking paradoxes, which is precisely why we posit variable ranking.

After discussing all of the relevant constraints, we propose the tentative ranking of all the constraints in (25), where *NUC/V represents the family of constraints ranked in accordance with the sonority hierarchy. As we have seen, this ranking is not always fixed and free rankings may occur across four major levels.

- (25) *NUC/C » {LINEARITY, NON-IDENTITY, LINEARITY-BD(G,V)} » PHONOTACT » {[ANCHOR(L,V) » MAX-IO(V) » MAX-IO(C)], [*FALLING » *RISING], *NUC/V}

We summarize the primary ranking in tableaux (26) and (27).

- (26) *u tsai + tiau k^hi > u tsau k^hi* ‘be able to go’

tsai + tiau	*Nuc/C ₂	LINEARITY	NON-IDENTITY	LINEARITY-BD(G,V)	PHONOTACT	ANCHOR(L,V)	MAX-IO(V)	MAX-IO(C)	*FALLING	*RISING	*NUC/u	*NUC/i	*NUC/a
tsaitiau	*!				*						*	**	**
tsaiiau					*!			*			*	**	**
tsaiu					*!		**	*			*	*	*
tsaui		*!			*		**	*			*	*	*
tsiau						*!	**	*			*	*	*
tsai			*!				***	*	*			*	*
☞ tsau							***	*	*		*		*
t _s a							****!	*					

(27) *ka*₂₁ + *gua ma* > *ka*₂₃ *ma* ‘scold me’

ka + gua	*NUC/C ₂	LINEARITY	NON-IDENTITY	LINEARITY-BD(G,V)	PHONOTACT	ANCHOR(L,V)	MAX-IO(V)	MAX-IO(C)	*FALLING	*RISING	*NUC/u	*Nuc/i	*NUC/a
kagua	*!				*						*		**
kaua					*!			*			*		**
kau				*!			*	*	*		*		*
kua						*!	*	*		*	*		*
☞ ka							**	*					*
ku						*!	**	*			*		

It is reasonable to ask at this point (and as noted by an anonymous reviewer) how the proposed OT analysis accounts for the existence of syllable contraction in the first place (note that such teleological questions never even arise in the context of derivational models). According to OT, the universally most unmarked output is silence, which of course violates no markedness constraint at all (not even *NUC/V). Like lenition generally, syllable contraction in Taiwan Southern Min approaches this ideal only partway, due to the conflicting demands of faith constraints (including perhaps quasi-morphological BD correspondence constraints). For fuller discussions of the phonetic forces motivating the particular markedness constraints that are ranked high in Taiwan Southern Min syllable contraction, see Tseng (1999) and Li (2005).

The question that arises now is whether GLA is able to learn this ranking as well. If not, we will need to determine if this is the fault of the algorithm or the fault of assumptions we have made about generalizations in the data.

4. THE GRADUAL LEARNING ALGORITHM

As noted in the introduction, the Gradual Learning Algorithm (GLA) is a fully automatic procedure for learning OT grammars from data; a computer implementation is available as part of the widely used Praat phonetic analysis software, available on the Web (Boersma and Weenick

2004). The theoretical interest of applying GLA here is the potential that the algorithm has for learning formal OT grammars even when the data involve variation. GLA builds on the fundamental proposal of Prince and Smolensky (2004) that an OT grammar consists of a number of ranked constraints, with every possible input (underlying forms) associated with a large number of output candidates and the single winning candidate output being determined by constraint ranking. However, in order to handle variability, GLA is grounded in a stochastic OT grammar.

We begin by explaining the notions underlying stochastic OT and GLA in section 4.1. Then in 4.2 we apply GLA to the full contraction data given in Hsu (2003), first by treating them as categorical (non-variable) and then including variable data. Next, in Section 4.3 we apply GLA to the partial contraction data collected in a production study described in Li (2005).

4.1 Stochastic OT

In the stochastic OT model in Boersma (1998) and Boersma and Hayes (2001), constraint ranking is not a simple relation of linear precedence (e.g. $A \gg B$ vs. $B \gg A$), but rather the ranking of a constraint in the hierarchy is associated with a continuous value. Thus if constraints A and B have the ranking relation $A \gg B$, this can be true in an infinite number of different ways: A and B may have, respectively, the values 10 and 9, 100 and -34, or 0.09 and 0.001. These continuous ranking values are assumed to form part of adult competence.

The only way in which these values are observable is in how they affect variability in performance. Namely, if two values are sufficiently close together, the associated constraints are more likely to be reranked in any given utterance, while if they are sufficiently far apart, the associated constraints will behave as if their ranking is fixed across utterances. Formally, the “sufficient” distance between values for causing or preventing variable ranking in performance is handled by a “noise” value representing the width of the range around the continuous value of any given constraint. This noise is assumed to be part of performance, not competence, and is thus identical for all constraints in the mature grammar. Thus the model assumes that speakers choose ranking values for each constraint at random from within the ranges; with highly overlapping ranges, reranking will be common, while with nonoverlapping ranges, reranking will be impossible. For purposes of mathematical elegance, the range around a constraint value is modeled as

a normal distribution (i.e. a bell curve) with the prototypical value of the constraint as the mean and the width of the distribution (noise value) represented with the standard deviation. The practical effect (as any introductory statistics textbook will tell you) is that about 68% of the area of the range is within one standard deviation of the constraint value, 95% is within two standard deviations, and over 99% is within three standard deviations.

To take a schematic example, suppose two constraints have the ranking values 100 and 10, respectively, with a noise value of 2. This means that the two constraints are 40 $(=(100-10)/2)$ standard deviations apart, so it is extremely unlikely that the two constraints will be reranked in performance. By contrast, if the noise value is 2 but the two ranking values are 100 and 99, respectively, there will be a notable probability of reranking. More generally, if two constraints have ranking values greater than two standard deviations apart (i.e. if the noise value is 2, the two constraint values are greater than 4 points apart), this means that the midpoint is one standard deviation from each ranking value. The above information about the area of a normal distribution thus implies that the probability of their reranking must be less than 32% $(=100-68\%)$. If the two constraint ranking values are more than four standard deviations apart, the probability of reranking is less than 5% $(=100-95\%)$, and if the distance is more than six standard deviations, the probability is less than 1% $(=100-99\%)$. In actual fact, when the probabilities are calculated properly (see formula in Boersma 1998:331), they are much, much lower. Boersma (1998:332) gives a table (repeated below in (28)) showing the predicted rate of reranking for two constraints whose ranking values have the indicated distances (assuming a noise value of 2).

(28) Probability (%) of reranking (after Boersma 1998:332)

Distance	0	1	2	3	4	5	6	7	8	9	10
Probability	50	36	24	14	7.9	3.9	1.7	0.7	0.2	0.07	0.02

Thus, as Boersma and Hayes (2001) note, even if two ranking values are merely five standard deviations apart (e.g. 100 vs. 90 with a noise value of 2), the probability of reranking is about 1/5000, which is so low as to be indistinguishable from a speech error. This means that stochastic OT is not only capable of describing variability, but also explaining why variability is not found in every phonological pattern.

The Gradual Learning Algorithm thus shows how a stochastic OT

grammar of this kind can be acquired from data. In essence, all GLA does is compare each actual data item with the output predicted by the grammar as hypothesized at that stage in development. Similar to the OT acquisition model of Tesar and Smolensky (2000), GLA posits that the constraints are innate and so need not be learned, and both models also utilize the simplifying assumption that the child already knows the input form and needs only to learn the proper ranking that will link it to the attested output form. If there is any mismatch between the predicted output form and the actual data item, the constraint values unique to the incorrect form will be demoted while those unique to the correct form will be promoted. The mechanics of this process are identical to those assumed by Tesar and Smolensky (2000) except that within its stochastic grammar framework, the demotions and promotions in GLA involve continuous constraint values rather than linear precedence. For example, in one learning cycle the constraint values for A and B may change from 98 and 92, respectively, to 100 and 90; thus they will still retain the same prototypical ranking, but the probability of their reranking in performance will be decreased. Given sufficient data, the GLA is able to perform the probability matching described in the introduction: whatever rate of appearance of alternate forms in the data, the mature stochastic grammar will generate outputs matching this rate.

While we do not claim that GLA will prove to be the “ultimate truth” as to how grammars are learned and structured, it does seem to be the best currently available model of how variability is learned. Keller and Asudeh (2002) indicate some problems they see with GLA, the most fundamental of which is its blurring of the line between competence and performance by the modeling of frequency distributions directly within the grammar. However, Keller’s own model of linguistic variability (Keller 2000, to appear) rejects OT premises that are far more fundamental than strict ranking, since it permits lower-ranked constraints to override higher-ranked constraints in certain cases. In addition, as a model of adult performance (specifically, grammaticality judgments), it does not propose any learning algorithm. Thus Keller’s model cannot provide an explanation of the acquisition of probability matching, whereas GLA can.

With this as background, we are now ready to turn to our applications of GLA to Taiwan Southern Min syllable contraction.

4.2 Full contraction

We began by testing GLA on the full contraction data from Hsu (2003), divided into two subgroups. The first subgroup contained only outputs which Hsu reported to be fully consistent with the Sonority model, which we will call the “categorical” data set. The second subgroup included all of the data listed in Hsu’s appendix (Hsu 2003:375), including alternative outputs, some of which deviated from the predictions of the model; we call this the “variable” data set. Given that our constraints and their basic ranking were primarily based on Hsu’s analysis of the first subgroup, we expected GLA to do very well with it, but the second subgroup better represents the variability of actual speech. In total there were 37 input-output pairs in the categorical data set and 49 in the variable data set (see Appendix 1). Each pair was hand-coded as to whether it obeyed or violated each of our proposed constraints.

The learning data for GLA consist of what are called pair distributions, where each input form is paired with each of its possible outputs with a weighting proportional to that in actual language data. Since Hsu (2003) does not provide frequency data, we otherwise assumed that alternative output forms appeared equally often. For example, the two outputs *sin* and *sun* in the pair *si + tsun > sin / sun* ‘moment’ were each assigned 50% of the distribution.

Following Boersma and Hayes (2001), we began the first training stage with a large value for noise and then dropped it down to the “adult” value of 2 (the actual value is arbitrary). We also followed them in gradually reducing what they call “plasticity,” which represents the amount by which the continuous constraint ranking values are adjusted (i.e. promoted and demoted). Use of reductions in values such as noise and plasticity is standard practice in the design of learning and search algorithms for the same reason that anyone searching for a small point in a vast space (e.g. a driver looking for a particular address or a lab technician focusing a microscope) will refine the precision of the search over time, either gradually or abruptly shifting from a coarse search to a fine search. Boersma and Hayes (2001) speculate that this may also reflect a genuine characteristic of child language acquisition: young children tend to be flexible and quick learners of phonology but as they improve in accuracy their learning also slows down until, as adults, it is difficult or impossible to learn any new (nonlexical) phonology.

The training schedule we used for both subgroups of Hsu’s data is shown in (29), following that used by Boersma and Hayes (2001:80) in their modeling of some data from Finnish. In every stage, an

input-output pair was randomly chosen around 1,000 times on average (the numbers varied slightly in each stage).

(29) Training schedule

Stages	Plasticity	Noise
First	2.0	10.0
Second	2.0	2.0
Third	0.2	2.0
Fourth	0.02	2.0
Last	0.002	2.0

GLA begins by assuming arbitrary ranking values for the innate constraints. Learning involves adjusting these rankings slightly with each new data input. Again following Boersma and Hayes (2001), in our applications of GLA the ranking value of each constraint in the initial stage was set with the arbitrary value of 100. The algorithm then compared the incoming learning data and adjusted the ranking values of all the constraints. If an incoming learning token violated some constraints, the algorithm demoted their rankings and promoted the rankings of others. This adjustment ensured that that the correct output would be more likely to be generated on any future occasion.

The mature grammar derived from the categorical data set is shown in (30). Higher values indicate higher ranking. The distance of the ranking values indicates the relative ranking relationship of the constraints. As a general rule of thumb, a between-constraint distance of 3.5 or less implies that alternative outputs are likely to be readily noticed (since the probability of reranking will go over 10%). Pairs of constraints with ranking values at least this close are indicated in the table by “R” (for “rerankable”) in the cells at the intersections of the relevant constraint names.

(30) Ranking values derived by GLA from the categorical data set

		Variable rankings															
Constraint	Ranking value	NON-IDENT	LINEARITY	*NUC/u	PHONOTACT	*NUC/C ₁	MAX-IO(V)	*NUC/C ₂	ANCHOR(L,V)	*NUC/o	LIN-BD(G,V)	*RISING	*Nuc/e	*Nuc/i	*Nuc/a	MAX-IO(C)	*NUC/ə
NON-IDENT	474.4																
LINEARITY	473.85	R															
*NUC/u	471.41	R	R														
PHONOTACT	470.4		R	R													
*NUC/C ₁	468.62				R	R											
MAX-IO(V)	468.5				R	R	R										
*NUC/C ₂	468.45				R	R	R	R									
ANCHOR(L,V)	466.2					R	R	R									
*NUC/o	465.33					R	R	R	R								
LIN-BD(G,V)	465.28					R	R	R	R	R							
*RISING	464.46								R	R	R						
*NUC/e	463.01								R	R	R	R					
*NUC/i	437.92																
*NUC/a	0.94																
MAX-IO(C)	-951.73																
*NUC/ɔ	-1856.47																
*FALLING	-3012.23																

The mature grammar derived from the variable data set is shown in (31), which uses the same conventions as in (30). In particular, note that as in (30), the “R” marks imply that rerankable constraints fall into rough blocks, corresponding to the “peaks” in the “R” pattern. For example, in (30) above, we could posit the block {*NUC/C₁, MAX-IO(V), *NUC/C₂}, corresponding to the second “R peak” (we do not include PHONOTACT in this block because it is included in the block above it).

(31) Ranking values derived by GLA from the variable data set

		Variable rankings																
Constraint	Ranking value	*Nuc/C ₂	*Nuc/C ₁	NON-IDENT	LINEARITY	LIN-BD(G,V)	*Nuc/u	PHONOTACT	MAX-IO(C)	MAX-IO(V)	*Nuc/o	*Nuc/e	ANCHOR(L,V)	*Nuc/i	*RISING	*Nuc/a	*Nuc/ɔ	
*Nuc/C ₂	156.94																	
*Nuc/C ₁	154.82	R																
NON-IDENT	152.76		R															
LINEARITY	149.02																	
LIN-BD(G,V)	118.65																	
*Nuc/u	34.58							R										
PHONOTACT	34.32							R	R									
MAX-IO(C)	33.43								R									
MAX-IO(V)	31.64							R	R	R								
*Nuc/o	30.07									R	R							
*Nuc/e	29.55										R	R						
ANCHOR(L,V)	29.28											R	R	R				
*Nuc/i	29.06												R	R	R			
*RISING	27.46														R	R	R	
*Nuc/a	27.21															R	R	R
*Nuc/ɔ	26.97																R	R
*FALLING	-4454.61																	

Treating “blocks” of constraints as explained above, we can schematize the two GLA-derived grammars in a format that allows for a somewhat clearer comparison with the “hand-derived” analysis of (25). These three analyses are shown in (32). Of course, as shown by the overlapping “R blocks” in (30) and (31), the constraint blocks in (32b-c) are not as strictly separated as those in (32a). Stochastic OT (hence GLA) is an inherently quantitative model, so it is the numerical values in (30) and (31) that determine how the model behaves in practice.

- (32) a. Hand-derived ranking [based on (6) and (25)]
 { *NUC/C₁, *NUC/C₂ } » { LINEARITY, NON-IDENTITY, LINEARITY-BD(G,V) } » PHONOTACT » { [ANCHOR(L,V) » MAX-IO(V) » MAX-IO(C)], [*FALLING » *RISING], [*NUC/u » *NUC/i » *NUC/o » *NUC/e » *NUC/ɔ » *NUC/a] }
- b. GLA-derived ranking based on categorical data [after (30)]
 { NON-IDENTITY, LINEARITY, *NUC/u, PHONOTACT } » { *NUC/C₁, MAX-IO(V), *NUC/C₂ } » { ANCHOR(L,V), *NUC/o, LINEARITY-BD(G,V) } » { *RISING, *NUC/e } » *NUC/i » *NUC/a » MAX-IO(C) » *NUC/ɔ » *FALLING
- c. GLA-derived ranking based on variable data [after (31)]
 { *NUC/C₁, *NUC/C₂, NON-IDENTITY } » LINEARITY » LINEARITY-BD(G,V) » { *NUC/u, PHONOTACT, MAX-IO(C), MAX-IO(V) } » { *NUC/o, *NUC/e, ANCHOR(L,V), *NUC/i } » { *RISING, *NUC/a, *NUC/ɔ } » *FALLING

Since the learning data were all fully contracted syllables in which intervocalic consonants were deleted, GLA ranked the constraints *NUC/C₂ and *NUC/C₁ at or near the top, which is particularly clear in the ranking derived from variable data in (32c). The similarly never-violated constraints NON-IDENTITY, LINEARITY, and LINEARITY-BD(G,V) are also ranked at the top, at least in (32c). Further, whereas PHONOTACT appears at the top in the GLA-derived ranking from categorical data in (32b), it appears more towards the middle in (32c), just as it does in our hand ranking in (32a).

The remaining constraints in (32c) that are ranked below PHONOTACT essentially form a cluster along the continual ranking scale (except for *FALLING, which we discuss below). This implies that these constraints were more easily reranked with respect to each other, just as is implied by the curly-brace notation in (32a). Considering the constraints in (32c) in the *NUC/V family, three sub-clusters are apparent: *NUC/u » { *NUC/o, *NUC/e, *NUC/i } » { *NUC/a, *NUC/ɔ }. This ranking is roughly consistent with the sonority hierarchy assumed by Hsu (2003).

One difference between (32c) and (32a) is that we expected the ranking MAX-IO(C) » MAX-IO(V) but instead found no evidence for ranking at all. This is not a flaw in GLA, but follows directly from the nature of our data sets. These two constraints never interact directly in

these data sets because the learning data involved only fully contracted syllables (i.e. each containing only one sonority peak); faithfulness to the intervocalic consonants was therefore irrelevant. We also expected ANCHOR(L,V) to outrank MAX-IO(V) in order to block the appearance of prevocalic glides intruding from the second syllable, but GLA did not induce this ranking from either data set. The fault here may lie with our constraints, which may not sufficiently distinguish surface glides that derive from the second syllable from those that derive from the first syllable. This mismatch thus highlights a practical benefit of GLA: it can help OT practitioners check whether their analyses actually fit the data.

A more serious problem is revealed by the constraint *FALLING, which was expected to be ranked above *RISING. Its appearance at the bottom of both GLA-derived rankings implies that it was instead entirely irrelevant. The reason for this was already anticipated in Section 3: the more general constraint ANCHOR(L,V) does the same job as *FALLING, as in *bo* + *iau kin* > *bua kin* ‘it doesn’t matter’. Note that in both GLA-derived rankings, ANCHOR(L,V) does indeed outrank *RISING. Strictly speaking, ANCHOR(L,V) and *FALLING are not in an “elsewhere” relation, but *FALLING is violated in precisely the same items where ANCHOR(L,V) is (at least in Hsu’s data set), whereas ANCHOR(L,V) also rules out other possible outputs (such as preservation of the vowel of the second syllable when it is not more sonorous). This again shows the usefulness of GLA in revealing a possibly redundant constraint.

Interestingly, as noted above, it seems that the rankings derived by GLA from the variable data set in (32c) more closely resemble the rankings derived in the “by-hand” analysis in (32a) than the rankings derived by GLA from the categorical data set in (32b). One possible explanation for this might be that the variable data set is a more accurate reflection of the actual pattern underlying the analysis of Hsu (2003), but this does not seem right, given that the additional items in the variable data set were problematic for this analysis. A more interesting possibility may be that learning a stochastic OT grammar benefits from being exposed to variable data. If this possibility is right, the finding might show more than simply a methodological advantage for the GLA, but may also reveal something about how actual human learners are able to cope with variable language data, and indeed why such variability is allowed to exist.

GLA can thus be said to have been mostly successful in inducing the correct grammar from contraction data, even when variability was involved. In the next section, we will test the model on variable data that

include partially contracted syllables and empirically derived frequencies.

4.3 Partial contraction

The partial contraction data used in our final GLA test came from a production experiment described in Li (2005). This experiment involved the “shadowing” (i.e. repeating back auditorily presented items) of 120 disyllabic words and phrases collected from the Taiwanese Spoken Corpus (Myers and Tsay 2003a), which consisted of a series of radio broadcast talk shows; these items are listed in Appendix 2.³ Since these items were originally chosen to test a different set of hypotheses, they did not overlap with those studied by Hsu (2003); in particular, they tended to be of lower frequency and hence were not contracted as often or completely. Note that these 120 items are likely to be more representative of fluent speech since they were chosen as a random sample, not to illustrate any particular phonological analysis.

Twenty college-aged Taiwan Southern Min native speakers living in southern Taiwan were asked to repeat the spoken items back naturally as soon as they heard them. The experimental procedure was performed using the DMDX experimental control software (Forster 2002) with the spoken responses recorded automatically. Without being explicitly told to do so, all speakers tended to contract the items to varying degrees. A total of 2,400 (= 120 × 20) recorded tokens were phonetically coded, with help of Praat (Boersma and Weenick 2004), as obeying or violating each of the constraints discussed above. This gave us a pair distribution of tokens reflecting estimates of the actual proportions of each alternate form in everyday speech.

The pair distributions were then input into GLA, which again started with all constraints set to an initial ranking value of 100. The training schedule was identical to that for the previous two GLA tests, with noise and plasticity decreased across the learning stages; each of the five learning stages consisted of approximately 185,000 input-output pairs.

³ An anonymous reviewer comments that a few items in Appendix 2 are not colloquial (i.e. items 88, 89, 95, 96, 105). Nevertheless, all were taken from our corpus of spontaneous speech, and their atypicality is in fact consistent with the goals of Li (2005), which focuses on lexical frequency effects (Appendix 2 lists items from highest to lowest frequency). Nevertheless, we admit that there may be a confound between frequency and other pragmatic factors (e.g. mainly spoken vs. mainly written). We plan to investigate the problem in statistical reanalyses of Li’s data, but the issue is not crucial here.

Modeling Variation in Syllable Contraction

The results of the mature grammar are shown in (33), with the same conventions as in (30) and (31) above.

(33) Ranking values derived by GLA from the partial contraction data

		Variable rankings															
Constraint	Ranking value	ANCHOR(L,V)															
		LINEARITY	NON-IDENT	MAX-IO(V)	*NUC/o	*NUC/ɔ	*NUC/i	*NUC/e	*RISING	LIN-BD(G,V)	*NUC/u	*FALLING	*NUC/a	PHONOTACT	MAX-IO(C)	*NUC/C ₁	
ANCHOR(L,V)	145.43																
LINEARITY	129.76																
NON-IDENT	104.77																
MAX-IO(V)	-280.62																
*NUC/o	-286.73																
*NUC/ɔ	-288.19					R											
*NUC/i	-288.59					R	R										
*NUC/e	-293.22																
*RISING	-550.40																
LIN-BD(G,V)	-724.60																
*NUC/u	-3560.39																
*FALLING	-4777.86																
*NUC/a	-6525.89																
PHONOTACT	-9587.83																
MAX-IO(C)	-9589.50														R		
*NUC/C ₁	-9592.08														R		
*NUC/C ₂	-9656.64															*NUC/C ₁	

With the same caveats as before, we can schematize the above ranking as in (34b), with the hand-derived ranking repeated in (34a) for comparison.

- (34) a. Hand-derived ranking [based on (6) and (25)]
 {*NUC/C₁, *NUC/C₂} » {LINEARITY, NON-IDENTITY, LINEARITY-BD(G,V)} » PHONOTACT » {[ANCHOR(L,V) » MAX-IO(V) » MAX-IO(C)], [*FALLING » *RISING], [*NUC/u » *NUC/i » *NUC/o » *NUC/e » *NUC/ɔ » *NUC/a]}

b. GLA-derived ranking based on partially contracted data

ANCHOR(L,V) » LINEARITY » NON-IDENTITY » MAX-IO(V) »
*NUC/o » {*NUC/ɔ, *NUC/i} » *NUC/e » *RISING »
LINEARITY-BD(G,V) » *NUC/u » *FALLING » *NUC/a »
PHONOTACT » {MAX-IO(C), *NUC/C₁} » *NUC/C₂

It is clear that the GLA-derived ranking in (34b), which is based on partially contracted data, is dramatically different from any of the rankings based on full contraction in (32). Most notably, the constraints *NUC/C₁ and *NUC/C₂ are now ranked at the bottom, whereas they appeared at the top with full contraction data. The reason for this is obvious: in most cases of partial contraction, intervocalic consonants are only reduced, not deleted, and thus violate *NUC/C (at least as we applied it, treating output forms as monosyllabic regardless of the degree of contraction). An observation of possible theoretical interest is that here the constraint *NUC/C₁ outranks *NUC/C₂, suggesting that the coda of the first syllable tended to drop off more easily than the onset of the second syllable. This presumably relates to the higher sonority of codas, on average, compared with onsets.

The constraints LINEARITY and NON-IDENTITY are ranked highest, over all other constraints, as was also the case for the previous analyses, though probably for different reasons; as an essentially phonetic process, we do not expect metatheses or complete neutralization. In particular, it seems implausible to assume that NON-IDENTITY was obeyed “on purpose” by the speakers; rather, it’s occurrence was merely an accidental side-effect of the phonetic nature of the partial contraction process. The hypothesis that the process is essentially phonetic is further supported by the high ranking of ANCHOR(L,V), which formerly appeared lower; speakers in this case seem to be following simple temporal order in preserving the first vowel in its original position. Apparently for the same reason, the constraint PHONOTACT lost most of its status; we do not expect a phonetic process to be structure-preserving. Note, however, that MAX-IO(V) is now ranked higher than MAX-IO(C), as we had originally expected, since vowels tend to resist the process of contraction better than intervocalic consonants. This contrasts with the modeling of the full contraction data, where MAX-IO(C) was simply irrelevant.

Given the plausibility of all as the above, it may seem surprising that GLA seemed to perform so badly in ranking all of the other constraints,

namely *FALLING, *RISING, LINEARITY-BD(G,V), and the *NUC/V family. However, the explanation of this seems quite simple: all of these constraints relate to syllable structure and hence are irrelevant unless a fully contracted syllable is produced. The ranking within the *NUC/V family, for instance, probably reflects more the accidental proportion of vowels across items in the data set than any sonority preferences, especially since without full contraction it is meaningless to talk about the “nucleus”.

The lesson here is that when dealing with a more “phoneticky” process, one should use constraints that reflect this, rather than testing constraints designed with categorical representations in mind. As it happens, the stochastic OT model proposed in Boersma (1988) is embedded in an approach that makes no significant distinction between phonology and phonetics (just as one might expect of the co-inventor of Praat); applications of Boersma’s phonetically detailed OT constraint formalism include Myers and Tsay (2003b). However, it would go far beyond the scope of the present paper to pursue this matter here.

5. CONCLUSION

In this paper we have shown how Taiwan Southern Min syllable contraction might be modeled in OT, emphasizing from the beginning that we must acknowledge some variability in the ranking. We then showed how GLA is capable of automatically inducing plausible rankings from different samples of data, revealing something, perhaps, about how children accomplish this task. Since constraint ranking in stochastic OT is seen as essentially continuous rather than discrete, even in competence, it is relatively straightforward to incorporate frequency information into OT grammars and thereby account for language variation. We hope that this is clear for at least one of the three types of variation mentioned in the introduction: multiple outputs for a single input. GLA may also be useful in dealing with the two other types, namely variability across items (e.g. higher frequency syllable sequences are more likely to contract than less common ones) and phonetic variation of the sort addressed in Section 4.3 (though to apply GLA properly we would also need phonetically detailed constraints). Stochastic constraint evaluation thus seems to be a promising mechanism in the construction of grammars that match language facts.

REFERENCES

- Alderete, John. 2001. *Morphologically Governed Accent in Optimality Theory*. New York: Routledge.
- Anttila, Arto. 1997. Deriving variation from grammar. *Variation, Change and Phonological Theory*, ed. by F. Hinskens, R. Van Hout, and W. L. Wetzels, 35-68. Amsterdam: John Benjamins.
- Apoussidou, Diana and Paul Boersma. 2004. Comparing two Optimality-Theoretic learning algorithms for Latin stress. *WCCFL* 23:29-42.
- Boersma, Paul. 1998. *Functional Phonology*. The Hague: Holland Academic Graphics.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45-86.
- Boersma, Paul and David, Weenick. 2004. *Praat: Doing Phonetics by Computer*. Retrieved July, 2005 from World Wide Web: <http://www.fon.hum.uva.nl/praat>.
- Cheng, Robert L. 1985. Sub-syllabic morphemes in Taiwanese. *Journal of Chinese Linguistics* 13 (1):12-43.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chung, Raung-fu. 1996. *The Segmental Phonology of Southern Min in Taiwan*. Taipei: The Crane Publishing Co.
- Chung, Raung-fu. 1997. Syllable contraction in Chinese. *Chinese Languages and Linguistics III: Morphology and Lexicon*, ed. by F. Tsao and S. Wang, 199-235. Taipei: Academia Sinica.
- Clements, George N. 1990. The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by John Kingston and Mary E. Beckman, 283-333. Cambridge: Cambridge University Press
- Forster, Jonathan. 2002. *DMDX Display Software*. Retrieved November, 2003 from World Wide Web: <http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>.
- Hsiao, Yuchau E. 2002. Tone contraction. *Proceedings of the Eighth International Symposium on Chinese Languages and Linguistics*, 1-16. Taipei: Academia Sinica.
- Hsu, Hui-chuan. 2003. A sonority model of syllable contraction in Taiwanese Southern Min. *Journal of East Asian Linguistics* 12 (4):349-377.
- Hsu, Hui-chuan. 2005. An Optimality-Theoretic analysis of syllable contraction in Cantonese. *Journal of Chinese Linguistics* 33 (1):114-139.
- Keller, Frank. 2000. *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. PhD dissertation, University of Edinburgh.
- Keller, Frank. To appear. Linear Optimality Theory as a model of gradience in grammar. *Gradience in Grammar: Generative Perspectives*, ed. by G. Fanselow, C. Féry, R. Vogel, and M. Schlesewsky. Oxford: Oxford University Press. Available at <http://roa.rutgers.edu>.
- Keller, Frank and Ash Asudeh. 2002. Probabilistic learning algorithms and Optimality Theory. *Linguistic Inquiry* 33:225-244.

Modeling Variation in Syllable Contraction

- Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.
- Li, Yingshing. 2005. *Frequency Effects in Taiwan Southern Min Syllable Contraction*. National Chung Cheng University MA thesis.
- McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics 18*, ed. by J. Beckman, L. Walsh Dickey, and S. Urbanczyk, 249-384.
- Myers, James and Jane Tsay. 2003a. *Phonological Competence by Analogy: Computer Modeling of Experimentally Elicited Judgements of Chinese Syllables (I)*. (Project report). National Cheng Chung University. Research Project funded by National Science Council, Taiwan. (NSC 91-2411-H-194-022).
- Myers, James. 2003b. A formal functional model of tone. *Language and Linguistics 4* (1):105-138.
- Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford, UK: Blackwell Publishing.
- Selkirk, Elisabeth O. 1984. On the major class features and syllable theory. *Language and Structure*, ed. by M. Aronoff and R. Oerhle, 107-136. Cambridge: MIT Press.
- Tesar, Bruce and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tseng, Chin-Chin. 1999. Contraction in Taiwanese: Synchronic analysis and its connection with diachronic change. *Chinese Languages and Linguistics V: Interactions in Languages*, ed. by Y.-M. Yin, L. Yang, and H. Chan, 205-232. Taipei: Academia Sinica.
- Venneman, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte 18*:1-18.
- Yip, Moira. 1988. Template morphology and the direction of association. *Natural Language and Linguistic Theory 6*:551-577.

Appendix 1.
Full contraction data (based on appendix in Hsu 2003:375)

	Input	Categorical output	Variable output	Gloss
1	bo + e	bue	bue / be	‘unable’
2	ho + gua	hua	hua	‘by me’
3	ka + gua ma	ka ma	ka ma	‘scold me’
4	tsa + k ^h i	tsai	tsai	‘morning’
5	loʔ + k ^h i	loi	loi	‘get down’
6	lai + k ^h i tɲ	lai tɲ	lai tɲ	‘go home’
7	bin + a tsai	mĩã tsai	mĩã tsai	‘tomorrow’
8	kim + a lit	kĩã lit	kĩã lit	‘today’
9	loʔ + ho t ^h ĩ	lo t ^h ĩ	lo t ^h ĩ	‘rainy day’
10	tsit + e	tse	tse	‘this one’
11	si + bo	sio	sio	‘right?’
12	toʔ + ui	toi	toi	‘where’
13	tsit + tsun	tsin	tsin	‘this moment’
14	hit + tsun	hin	hin	‘that moment’
15	li + k ^h ũã	nĩã	nĩã	‘look!’
16	u tsai + tiau k ^h i	u tsau k ^h i	u tsau k ^h i	‘able to go’
17	li + tsap	liap	liap	‘twenty’
18	e hiau + t ^h aŋ k ^h i	e hiaŋ k ^h i	e hiaŋ k ^h i	‘know how to go’
19	tsa + hɲ	tsaŋ	tsaŋ	‘yesterday’
20	m + t ^h aŋ	baŋ	baŋ	‘can not’
21	sã + tsap si	sãm si	sãm si	‘thirty-four’
22	tũĩ + lai	tuai	tuai	‘come back’
23	to + lai	tuai	tuai	‘come back’
24	loʔ + lai	luai	luai	‘fall down’

Modeling Variation in Syllable Contraction

25	bo + iau kin	bua kin	bua / bau kin	‘it doesn’t matter’
26	na + e an ne	nai an ne	nai an ne	‘how come?’
27	tso + bɔ laŋ	tsau laŋ	tsau / tsɔ laŋ	‘woman’
28	sio + kaŋ	siaŋ	siaŋ	‘the same’
29	u te + t ^h aŋ k ^h i	u taŋ k ^h i	u taŋ k ^h i	‘have somewhere to go’
30	he + ɔ	hiɔ	hiɔ	‘interjection for a sudden realization’
31	hɔ + guan	huan	huan	‘by us (exclusive)’
32	tsia + e	tsiai	tsiai	‘these’
33	hia + e	hiai	hiai	‘those’
34	lip + lai	liai	liai	‘come in’
35	k ^h i + lai	k ^h iai	k ^h iai / k ^h ai	‘get up’
36	ke + lai	kiai	kiai / kai	‘come over’
37	si + tsun	sin	sin / sun	‘moment’
38	hɔ + laŋ		hɔŋ / haŋ	‘by someone’
39	tsia? + nĩ		tsian	‘this’
40	hia? + nĩ		hian	‘that’
41	sia + mĩ laŋ		siam / sĩa laŋ	‘who’

Appendix 2. Partial contraction data

	Input	Gloss
1	piŋ + iu	‘friend’
2	hit + le	‘that one’
3	kam + kak	‘feel’
4	e + sai	‘able’
5	tsai + ã	‘know’

6	tak + ke	‘everyone’
7	k ^h o + liŋ	‘maybe’
8	ka + ti	‘oneself’
9	lai + te	‘inner’
10	mi? + kĩa	‘stuff’
11	i + kiŋ	‘already’
12	bun + te	‘question’

13	k ^h ui + ts ^h ia	‘drive a car’
14	si + kan	‘time’
15	tɔŋ + zen	‘for sure’
16	tai + uan	‘Taiwan’
17	iŋ + kai	‘should’
18	tset + bɔk	‘program’
19	loʔ + k ^h i	‘fall down’
20	ho + tsiaʔ	‘delicious’
21	tien + ue	‘telephone’
22	tu + tsiaʔ	‘just now’
23	k ^h i + lai	‘get up’
24	tai + tsi	‘thing’
25	p ^h ɔ + t ^h ɔŋ	‘ordinary’
26	m + ko	‘but’
27	i + au	‘after’
28	kuan + he	‘relation’
29	kue + bin	‘allergy’
30	sin + t ^h e	‘body’
31	kok + ui	‘everyone’
32	kɔŋ + ue	‘talk’
33	kan + tan	‘simple’
34	pe + bu	‘parents’
35	tsun + pi	‘prepare’
36	tɔŋ + iau	‘important’
37	ka + gua	‘to me’
38	un + tɔŋ	‘exercise’

39	u + kau	‘enough’
40	na + e	‘how’
41	to + ui	‘where’
42	iŋ + gi	‘English’
43	bak + kīã	‘glasses’
44	t ^h au + ke	‘boss’
45	ts ^h iŋ + k ^h i	‘clean’
46	to + sia	‘thank’
47	hūã + hi	‘happy’
48	ts ^h an + t ^h īã	‘restaurant’
49	be + hiau	‘unable’
50	tau + iu	‘soy-bean sauce’
51	hɔ + laŋ	‘by someone’
52	ts ^h in + ts ^h ai	‘casual’
53	tai + hak	‘university’
54	ket + hun	‘marry’
55	loʔ + hɔ	‘rain’
56	kɔŋ + kue	‘have talked’
57	tien + nau	‘computer’
58	iu + iŋ	‘swim’
59	tsa + k ^h i	‘morning’
60	hɔ + siɔŋ	‘each other’
61	tsi + u	‘only’
62	tsɔŋ + kiɔŋ	‘total’
63	gan + k ^h o	‘ophthalmology’
64	ien + tsau	‘play an instrument’

Modeling Variation in Syllable Contraction

65	tsiu + ni	‘anniversary’
66	sui + si	‘anytime’
67	tai + siŋ	‘beforehand’
68	k ^h au + tsai	‘eloquence’
69	to + ien	‘director’
70	bin + kan	‘folk’
71	ge + sut	‘art’
72	təŋ + tsək	‘action’
73	tsi + tsio	‘at least’
74	ləŋ + tioʔ	‘collide with’
75	taŋ + tse	‘together’
76	bi + sɔ	‘gourmet powder’
77	p ^h ũã + tuan	‘judge’
78	ki + p ^h io	‘airplane ticket’
79	be + hu	‘there is not enough time (to do something)’
80	tsi + ha	‘below’
81	tsui + tsun	‘standard’
82	hap + ts ^h ũ	‘chorus’
83	hueʔ + ap	‘blood pressure’
84	ki + kan	‘period’
85	pit + iau	‘necessary’
86	bi + iəŋ	‘cosmetology’
87	iu + lam	‘sightseeing’
88	k ^h i + ts ^h uan	‘asthma’
89	k ^h o + si	‘but’

90	tsiu + giap	‘get a job’
91	zin + tsai	‘talent’
92	k ^h i + hau	‘climate’
93	ki + kim	‘fund’
94	se + zi	‘careful’
95	huan + tsij	‘anyway’
96	kiŋ + zien	‘unexpectedly’
97	te + kiu	‘earth’
98	iəŋ + kam	‘brave’
99	sat + siŋ	‘kill’
100	tioʔ + kip	‘worry’
101	tək + p ^h in	‘drug’
102	ti + iu	‘lard’
103	tsik + zim	‘duty’
104	tsiaʔ + tiau	‘eat up’
105	tsij + kiŋ	‘ever once’
106	piŋ + siəŋ	‘ordinary’
107	bu + to	‘dance’
108	tsu + ts ^h eʔ	‘register (at school)’
109	kua + ho	‘register (in hospital)’
110	t ^h e + keʔ	‘physique’
111	liəŋ + sim	‘conscience’
112	ti + an	‘public security’
113	u + ziam	‘pollution’
114	tsiəŋ + kin	‘nearly’
115	pue + au	‘at the back’

Yingshing Li; James Myers

116	tsəŋ + kau	‘religion’
117	gan + kəŋ	‘eyesight’
118	ho + pit	‘why bother’

119	ts ^h i + k ^h u	‘urban district’
120	tai + k ^h uan	‘loan’

Yingshing Li
Graduate Institute of Linguistics
National Cheng Chung University
Ming-Hsiung, Chia-Yi, Taiwan
g9115007@ccu.edu.tw

James Myers
Graduate Institute of Linguistics
National Cheng Chung University
Ming-Hsiung, Chia-Yi, Taiwan
lngmyers@ccu.edu.tw

台灣閩南語音節連併變化之模型建構

李盈興、麥傑
國立中正大學

本文嘗試運用一種推測性優選理論模型，即為漸進學習演算法(the Gradual Learning Algorithm)，來建構台灣閩南語音節連併的變化。為探測此模型的有效性，三種複雜度互異的資料因而置入模型：第一種類型為符合許(2003)分析的完全連併音節；第二種類型為除了前項類型之外，增加了許(2003)分析的特例，但其仍為以閩南語為母語者認可的完全連併音節；第三種類型取自李(2005)語言發聲實驗的半連併音節，最為突顯語音的變化。結果顯示此模型能夠在置入不同資料後，分別提供合理的制約排列順序，同時能夠涵括其一般性及變化性。因此，本文證實推測性優選理論模型似乎能夠建構符合語言事實的語法。