THAO PHONOLOGY

Paul Jen-kuei Li

1. Introduction[1]

Even though Thao is virtually on the verge of extinction, very little work has been done on the language. Ogawa and Asai (1935:4) only mentioned about it in their monumental work, but made no field study of it. Abe (1930:430-35), Alvarez (1915-24), Bullock (1874-5), Collingwood (1868), and Steere (1874) have only short lists of vocabulary.

The best study of the language to date is Fang Kuei Li et al's (1956:23-51) "Notes on the Thao Language," which gives a brief description of the phonology and morphology. Its phonetic transcription is very accurate and its grammatical observation is insightful, regardless of the investigators' short stay at the village (only five half days). Its some 800 lexical items have become the main source of Thao for comparative Austronesian studies, e.g., Dyen (1963, 1965, 1971), Dahl (1973), Tsuchida (1975). However, the data collected were severely limited. Only two short texts with both Chinese and English translations ran only a little more than two pages. It is, therefore, essential to collect more linguistic data for this soon dying-out language.

I spent three weeks (June 24-July 5, July 12-20, 1975) at

---

1. This is one of the field reports of the Project on Formosan Language Survey, supported by the National Science Council, Republic of China. Pang-hsin Ting has read an earlier version of this paper and given valuable suggestions for improvement, and so has Prof. Fang Kuei Li.

Sun-Moon Lake, working on the language.[1] The bulk of my work was collection of texts and vocabulary, as well as some sentences elicited for the purpose of syntactic analysis. So far I have got 26 short texts (75 typewritten pages with English translation) and about 1,500 lexical items.

My main informant was Ch'iu-hsiang Mao 毛秋香, a 51 year-old woman. She was born of a Taiwanese father and a Thao mother, and was adopted by a Thao family when she was only 40 days old. She was brought up speaking Thao and did not learn to speak Taiwanese until she was ten years old when her family with others moved from Lalu 光華島 to Barawbaw 卜吉社, now called Te-hua She 德化社, as the construction of the Lake was complete (see Text XXI). Her 53 year-old husband rawʃian (Lao-hsien Mao 毛老先 in Chinese) also served as my informant from time to time. They both speak excellent Taiwanese, in which the informant sessions were conducted. In addition, they know some Japanese, Mandarin and a little Bunun. In the daily conversation, the Thao people speak much more Taiwanese than Thao. Most young people in the village do not speak Thao, although some of them have some passive understanding of the language.

One purpose for writing this article is to make the study of Thao phonology available to more readers. Prof. Li's (1956) report was written in Chinese although English glosses were also given for the lexical material. Consequently readers without knowledge of Chinese have no more access to Thao materials than

---

1. Prof. Fang Kuei Li joined me in two informant sessions and offered some very valuable suggestions. The influence of his paper on my study is apparent throughout this report.

the lexical.

This study generally follows the sound system as presented in Prof. Li but differs from it by setting up the phonemic contrast between /s/ and /θ/ and by eliminating one vowel phoneme /e/, as evidenced in the Thao language when more data are collected. These revisions are justified from synchronic as well as diachronic point of view. Furthermore, the generative approach to Thao phonology adopted in this study differs from Prof. Li's structural approach. Thus the phonology accounts for the problems of morphophonemic alternations that were originally treated in the section of morphology by him. None the less, I owe a great deal to him for all his insightful and enlightening observations of the language.

2. Phonology

2.1. Consonants and Vowels

Table 1. Consonants and Vowels

Consonant Phonemes:

		Bilab.	Dental	Palatal	Velar	Uvular	Glottal
Stops	vl.	p	t		k	q	ʔ
	vd.	b	d				
Nasal	vd.	m	n		ŋ		
Fric.	vl.	f	θ, s	ʃ			h
	vd.		đ				
Semi.	vd.	w		y			
Lat.	vl.		ɬ				
	vd.		l				
Trill	vd.		r				

Vowel Phonemes: i u

a

Table 2. Distinctive as well as Non-distinctive Features of Thao Segments

	i	u	a	y	[β] w	r	l	ɬ	p	[ʔb] b	f	m	t	[ʔd] d	θ	ð	n	s	ʃ	k	ŋ	q	ʔ	h
syll	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cons	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
high	+	+	-	(+)	(+)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	-	+	+	-	-	-
back	-	+	(-)	-	+	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	+	(+)	+	-	(-)
ant	(-)	(-)	(-)	(-)	(-)	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-
cor	(-)	(-)	(-)	(-)	(-)	(+)	(+)	(+)	-	-	-	-	+	+	+	+	+	+	(+)	(-)	(-)	(-)	-	-
voice	(+)	(+)	(+)	(+)	(+)	(+)	+	-	-	+	(-)	(+)	-	+	-	+	(+)	(-)	(-)	(-)	(+)	(-)	(-)	(-)
cont	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	-	-	+	-	-	-	+	+	-	+	+	(-)	(-)	(-)	-	+
nasal	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	-	-	(-)	+	-	-	(-)	(-)	+	(-)	(-)	-	+	(-)	(-)	(-)
strid	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	-	-	(-)	+	(+)	(-)	(-)	(-)	(-)	(-)

*Redundant features (or non-distinctive features) are put in parentheses.

2.2. Phonetic Descriptions

/p, t, k, q, ʔ/ are voiceless unaspirate stops,[1] of which /t/ is a very fronted dental stop, with the tip of tongue touching the back of teeth or even protruding between the upper and lower teeth. /k/ contrasts with /q/, e.g., paŋka 'chair,' paŋqa 'to rest.' /b, d/ are preglottalized voiced stops [ʔb, ʔd] respectively and sound like implosives. As noted in Li (1956:23), they are treated as single phonemes because [b] and [d] never appear without the preceding [ʔ] in the language. They cannot be treated as phonetic variants of /ʔp, ʔt/ because there are no parallel combinations such as /ʔk, ʔq/. Note that there is no /g/ [ʔg]; it is quite common to have the phonetic gap, cf. Tsou (Tung 1964, Tsuchida 1975), Vietnamese, and the South Min dialect of Hainan.[2]

A stop consonant between syllables is transitional and phonetically present in both syllables, e.g., /ta:ta/ [tá:t'taʔ] 'one,' /θu:may/ [θú:m'may] 'bear.' (The symbol ' indicates syllable division).

/f/ [Φ] is a bilabial fricative, not labiodental as the symbol indicates. /θ, đ/ are interdental fricatives. /đ/ [ð] sounds somewhat like [l] in word-final. In fact, the final đ of some items, e.g., fu:rađ 'moon,' qiɬpuđ 'star,' ti:ɬađ 'sun,' was mistranscribed as l in the early stage of the field work and

---

1. Stops immediately followed by h sound like aspirate stops; but interpreted as consonant clusters, e.g., puthu:m 'gun,' pha:laq 'to hunt deer,' since the structure of the language allows consonant clustering in both initial and medial positions; see §2.5 below.

2. Prof. F. K. Li (private conversation) has an explanation for the gap: It is physiologically more difficult to produce [ʔg] by having both glottal and voiced velar closures almost simultaneously.

corrected only after carefully checking up with the informants.[1] /s/ is a very fronted dental voiceless fricative in contrast with the interdental /θ/, e.g., ma:say 'jealous' vs. ma:θay 'dead,' sa:ɖum 'water' vs. θaw 'person,' sin?an 'to drink' vs. θi:θu 'he.' The distinction between /s/ and /θ/ is made by all the speakers of Thao that I observed, including Ch'iu-hsiang Mao, Rawshian and Pita.[2] /ʃ/ is a voiceless alveopalatal fricative, similar to English "sh," but pronounced with lips a little more rounded, as observed in Li (1956:23).

The sounds [β] and [w] are in complementary distribution, i.e., [β] in syllable-initial stressed position and [w] in syllable-medial and -final positions, e.g., [βákβak] 'jaw,' [táwn] 'house,' [θáw] 'person.' Although [w] may occur in a syllable-initial but unstressed position, e.g., [wa] 'possessive particle,' it is still in complementary distribution with the weak stressed [βa] as in [βaɖá:qan] 'lake.' Thus the two phonetically similar sounds are assignable to the same phoneme. Nevertheless, there is a problem of its position in the phonemic chart: As a voiced counterpart of the bilabial fricative (v vs. f), or as a labiovelar counterpart

---

1. The three items as given in Li (1956:31) are fú·ral 'moon,' kíɬpul 'star,' tí·ɬal 'sun.' The final -l is probably a mistranscription of ɖ for all these items. (Prof. Li's ɣ is equivalent to my ɖ.) PAN *N has regularly derived as Thao ɖ, e.g., *DaNum > sa:ɖum 'water' (θá·ɣum in Li 1956:31), *quZaN > qu:saɖ 'rain' (qú·θaɣ in Li 1956:31), *bulaN > fu:raɖ 'moon,' *tuNa > tu:ɖa 'eel.' The -l in Prof. Li's fú·ral, for instance, would be an unexplainable and irregular reflex of PAN *N.

2. Thao s and θ are derived historically from different sources: PAN *d, *D and *Z > Thao s, e.g., *laHud > mana-ráws 'downhill,' *DaNum > sa:ɖum 'water,' *quZaN > qu:saɖ 'rain,' *Zalan > sa:ran 'road,' and PAN *C > Thao θ, e.g., *Caŋis > θa:nit 'weep!' *Caqi > θa:qi 'excrement,' *Cau > θaw 'person,' *kuCuH > ku:θu 'head louse.'

of the alveopalatal semivowel (w vs. y)? Whichever position it should take, there would be a phonetic gap. The choice of the symbol w rather than y in this study is partly due to historical consideration.[1] If w is treated as the base and β as derived, then the rule can be stated as follows:

$$\text{R1.}\quad \begin{bmatrix}-\text{syll}\\-\text{cons}\\+\text{back}\end{bmatrix} \dashrightarrow \begin{bmatrix}+\text{cons}\\-\text{back}\\+\text{ant}\\-\text{cor}\end{bmatrix} \Big/ \underline{\quad} \begin{bmatrix}+\text{syll}\\+\text{stress}\end{bmatrix}$$

The above rule applies when preceding a vowel, but excludes the unstressed particle such as wa. Note that [+stress] indicates any degree of stress: primary, secondary, tertiary, including the weak stress.

Incidentally, the phonetic stress assignment rule (R2; see §2.3 below) must apply before the spirantizing rule (R1) above.

It is more economical to treat the glottal stop ʔ in word-final as non-phonemic. Like the South Min dialects of Chinese, the glottal stop in word-final is phonetically present but phonemically insignificant. It is present when a lexical item is uttered in isolation, but disappears when followed by a suffix or even by other free words in an utterance, e.g., matanwa:raʔ 'to choose,' ʔa-matanwara-yɖaʔ 'will choose;' θi:θuʔ 'his,' θi:θu wa ʔi:naʔ 'his mother,' θi:θu wa ʔi:na ma:ra sa:ɖum 'his mother took water.' Furthermore, the word-final glottal stop does not behave like a consonant. For example, the verb stems

1. See Li 1974 for the historical implications of the alternations between semivowels (or semiconsonants) and fricatives (or liquids). The Thao phoneme under discussion is historically derived from PAN *w, e.g., *babaw > fa:faw 'up,' *wiRiH > tana-wi:ɬi 'left,' *quway > qu:way 'rattan,' or from PAN *u under special developments, e.g., *laHud > mana-raws 'downhill.'

that end with a consonant take the suffix -in in the passive construction, e.g., ka:n-in 'to be eaten,' ʔyu:p-in 'to be blown,' θanu:p-in 'to be buried,' but words that phonemically end with a vowel take only -n, e.g., /ka:θu/ [ká:θuʔ] 'to bring' > ka:θu-n 'to be brought,' /minfa:ri/ [minɸa:riʔ] '(the wind) to blow,' > ʔinfa:ri-n 'to be blown away (by the wind).'

Alternatively, if the word-final ʔ is phonemically represented, then the problem is that it does not behave like a true consonant only in this position and would have to be treated as different from all the other consonants, including the semivowel /w/ (see below).

The glottal stop in the word-initial presents more problem than in the word-final. It disappears when prefixed in the item: ʔi-nay 'here' > ʔa-y-nay 'will be here' (ʔa- is the 'future' marker), but preserved in many items such as: ʔaɖa:ɖak 'child' > maɬi-ʔaɖa:ɖak 'to have a baby,' ʔa:puy > ɬmin-ʔa:puy 'to produce fire.' It is, therefore, necessary to represent the glottal stop in the phonemic representation in both word-initial and medial positions.

Judging from the suffixation, incidentally, word-final w behaves like a true consonant, e.g., [mákʃnaw] 'to breathe,' [pakʃná:β-in] 'to be breathed (?);' [ɸá:riw] 'to buy/sell,' [ɸariβ-í:ɖa] 'has bought,' [ɸarí:β-in] 'to be bought/sold,' [ɸarí:β-i] 'sell it!' [ʃá:ʃaw] 'to chase away,' [ʃaʃá:β-in] 'to be chased away.'

In summary, there are altogether 21 consonants /p, t, k, q, ʔ, b, d, m, n, ŋ, f, θ, ɖ, s, ʃ, h, l, ɬ, r, w, y/ in Thao. The list of consonants in this study differs from Li et al's (1956)

by adding the phoneme /s/. In addition, single symbols are used to stand for single phonemes, i.e. /b, d/ rather than /ʔb, ʔd/.

Thao has only the three main vowels /a, i, u/. Both high vowels are fairly open. /i/ has the variant [e] next to /h, ʔ, r/, e.g., /ri:buʃ/ [ré:buʃ] 'bush,' /ba:hi/ [ʔbá:heʔ] 'soul.' When /i/ occurs next to /q/, there is a transitional [ə] in between, e.g., miqi:ɬa [miəqəí:ɬaʔ] 'to drink (wine).' /u/ has the variant [o] or even [ɔ], particularly next to /q,r/, e.g., /pu:nuq/ [pú:nɔq] 'head,' /ku:pur/ [kú:pɔr] 'body hair,' /qaru:ta/ [qarɔ́:taʔ] 'cat.'

Prof. Li (1956:24) recorded a fourth vowel /e/ and noted, "In fact e is rare and usually occurs next to q and r," e.g., qmé·rqer 'to bite,' tapé·riʔ 'wild cat,' as appeared in his word list. However, the vowel does not contrast with any of the three vowels mentioned above. The vowel e in the cited examples are conditioned by r. Hence, it can be treated as a variant of /i/.[1]

2.3. Stress

(1) Stress generally falls on the penult, e.g., qú:sad 'rain,' ʃiráwsin 'rainstorm,' munrú:da 'to row a boat.' However, stress

---

1. Historically, the PAN, Proto-Hesperonesian (PHN) and Proto-Southern Formosan (PSF) (All these Proto- forms are based on Tsuchida 1975, but with the numerical subscripts left out) *ə has mostly derived as Thao i, occasionally as u by assimilation to the following vowel, or simply dropped out, e.g., PSF *DakəS > ʃa:kiʃ (ʃ- instead of anticipated -s is probably assimilated to -ʃ) 'camphor laurel,' PSF *bukəS 'hair' > fu:kiʃ 'hair,' PHN *salən > ta:rin 'pine-tree,' PHN *Sajək > ʃa:dik 'smell,' PHN *tapəS > ta:pis 'winnow,' PHN *Zəmaq > si:maq 'tomorrow;' PAN *təluH > tu:ru 'three,' PAN *təbuS > tufu:iʃ 'sugar cane' (The third vowel i is unexplained,) PHN *QaSəluH > qaʃu:ru 'pestle,' PHN *qə(n)tut > q/un/tut 'break wind' (cf. qtu:t 'fart'); PAN *kaʔən > ma-ka:n 'eat,' PFN *QəluD > qru:s 'pillar,' PHN *tənəʔun > tunú:n 'weave,' PAN *xəpat > ʃpa:t 'four,' PAN *baqəRuH > faqɬu 'new.'

falls unpredictably on the final syllable in a few items, e.g., ʃmiráws, 'flood, typhoon' (cf. ʃiráws-in above), ʔaɗyá:ɗ 'termite,' tampapuráy 'evil spirit,' qadadáw 'dung beetle,' pyaθawθáwn 'relative,' tunú:n 'to weave.' Perhaps some of these final stressed syllables are stems (see (2) below).

(2) Stress never falls on a prefix. If a penultimate syllable is a prefix or reduplication, then stress falls on the final syllable which is the stem, e.g., mu-táwn 'to go home,' ma-ká:n 'to eat,' mu-náy 'to come,' ʔi-náy 'here,' ʃaw-na-náy 'to arrive here,' ʔa-na-náy 'bring here,' tana-dú: 'right (side),' ʔi-dáy 'this.' There are only a few lexical items of this type.

(3) Stress may, however, fall on a suffix, e.g., ʔa-ma-kan-í:ɗa 'will be eating.' In such case, stress is still on the penult.

(4) Monosyllabic particles are unstressed, e.g., tu, ti, sa, qa, wa, ya, etc.

(5) The position of stress may vary from speaker to speaker, or even with the same speaker. For example, Prof. Li (1956:33) recorded the term for 'rib' as fá·ɬan, whereas I recorded it as faɬá:n. He (Li 1956:38) recorded pitáw, whereas I recorded pítaw for the same term 'door.' As a matter of fact, he also recorded pítaw elsewhere (§2.502). I have also recorded stress on either the penultimate or final syllable for the same items with the same informant.

Since stress mostly falls on the penult, and only a few items (including the exceptions mentioned in (1) and the prefixed monosyllabic stems illustrated in (2) above) receive stress on the final, stress is mostly predictable and will be marked only for the exceptions, i.e. the ones that receive stress on the final

syllable unpredictably, in this study. This solution departs from the principle of "once a phoneme, always a phoneme" in autonomous phonemics, but has the advantage of economy as advocated by systematic phonemics.

The predictable stress can be stated by the stress placement rule:

R2. [+syll] --→ [+stress]/___([-syll])([+cons])([+syll])([-syll])([+cons])#

Conditions: The [+syll] on the left of the arrow (1) does not contain a prefix and (2) is not a monosyllabic particle.

2.4. Length

(1) Stressed vowels are generally long and unstressed vowels are short, e.g., minɬá:fut 'sibling,' θmá:nit 'to cry,' ʃpú:t 'Chinese.' But there are exceptions to the stressed vowels, e.g., máta∫ 'to write,' ʔáʔa 'baby,' maráʔin 'big.'

(2) A stressed vowel immediately followed by two or more consonants is generally short, e.g., páŋqa 'table, chair,' ríθkuy 'orange,' táqryu 'gourd.' Exceptions: pí:ɬnaḍ 'lightning,' má:kθin 'ten,' masá:lpu 'worried.'

(3) A falling diphthong such as aw and ay is always short, e.g., θaw 'person,' tawn 'house,' barawbaw 'village name,' θayθuy 'they,' baylu 'beans,' θuyni 'today, now,' unless the semivowel is immediately followed by a vowel, e.g., fari:wi 'sell!'

In short, vowel length is mostly predictable: it generally coincides with stress. A stressed vowel is generally long if not followed by a semivowel alone or consonant cluster. To state positively, a stressed vowel is long before a true consonant,

which may or may not be followed by a vowel, or before a semivowel followed by a vowel:

R3. [+stress] --> [+long] / ___ { [+cons] ([+syll]) / [-syll] [+syll] [-cons] }

2.5. Distribution

All consonants except ŋ occur word-initially.

All consonants except b, d and h occur word-finally.

All consonants except ŋ occur intervocalically.

All consonants except ŋ may follow another consonant.

The consonant ŋ is rare and usually occurs before q or k,[1] e.g., miɬuŋqu 'to sit,' paŋqa 'to rest,' taŋqa:pu 'sister's child (?),' ɬuŋqawʃin 'to sneeze,' ʔiŋqa 'don't,' makaliŋkin 'healthy, strong,' paŋka 'chair.' It very rarely precedes other consonants, e.g., ʃaŋlaw 'vegetable.' It also rarely occurs word-finally, e.g., piʃtiŋtiŋ 'to be upset,' nuŋnuŋ 'end of spine,' tayʔiŋ 'European.'

In short, ŋ never occurs syllable- or word-initially, intervocalically, or after another consonant. The severe restriction of the distribution of the consonant leads me to suspect that it may not have a full phonemic status;[2] it is mostly a variant of m or n, but assimilated to the following consonants.

---

1. Prof. Li (1956:35) recorded the form ɬuŋpí·jaq 'twins,' whereas I recorded ɬumpi:yaq for the term. (His · is equivalent to my : and nis j my y.)

2. Historically speaking, PAN, Proto-Southern Formosan (PSF), and Proto-Hesperonesian (PHN) *ŋ has all derived as Thao n, e.g., PAN *Caŋis > θa:nit 'weep!' PAN *laŋaw > ranaw 'a fly,' *CaliŋaH > ɬari:na 'ear' (the initial ɬ instead of the anticipated θ is unexplained); PSF *ba[rR]aŋ > faɬa:n 'ribs;' PHN *saləŋ > tarin 'pine tree.'

In other words, there is no ŋ in the underlying representations especially for all the ŋ's immediately followed by q or k. The nasal assimilation rule can be stated as follows:

R4. $\begin{bmatrix}+\text{nas}\end{bmatrix} \dashrightarrow \begin{bmatrix}+\text{back}\end{bmatrix} / \_\_\_ \begin{bmatrix}-\text{cont} \\ +\text{back}\end{bmatrix}$

The rule above states that a nasal must be velar preceding a back stop. However, a nasal may or may not be assimilated to the following front stop, e.g., humbá:q 'ox,' bantaθ 'thigh,' barundud 'precisely,' but dumdum 'evening.'

The following statements are based on Prof. Li (1956:24):[1] "There are numerous combinations of two consonants,

(1) in the word-initial position, such as qm-, qn-, qt-, qc-, qf-, qθ-, qɬ-, ʃp-, ʃt-, ʃk-, ʃm-, ʃn-, ʃr-, ɬm-, ɬt-, θm-, θk-, hm-, km-, tm-, pr-, kr-, mr-, kɬ-, fɬ-, ph-, py-, my-, etc.;

(2) in the word-final position, only a semivowel followed by a consonant is possible, such as -wn, -wθ, -wd, -wɬ, -wl, -wk, -ys, -yn, etc., other consonant clusters do not appear word-finally;

(3) in the word-medial position, i.e. consonant clusters in the intervocalic position, there are too many consonant clusters and too many kinds to be enumerated...

Combinations of three consonants occasionally appear in the word-medial, such as -pny-, -ŋkm-, -kʃn-, nʃr-, ypɬ-, etc."

The statements (1) and (2) above can be revised as below, based on my own data:

(1) There are the following combinations of two consonants in the

---

1. I have changed some of Prof. Li's symbols to conform to my system, e.g., ʔb to b, ʔd to d, j to y, x̌ to đ, s to ʃ, etc.

word-initial position: pɖ-, ps-, ph-, pr-, py-; tm-; km-, kɖ-, kɬ-, kl-, kr-; qp-, qt-, qd-, qm-, qn-, qf-, qθ-, qɬ-, qr-; ʔy-; mr-, my-; fɬ-; θk-, θm-, θy-; sʔ-, sm-, sɖ-; ʃp-, ʃk-, ʃq-, ʃm-, ʃn-, ʃr-; ɬt-, ɬq-, ɬm-. Of these combinations, pɖ-, ps-, kɖ-, kl-, qp-, qr-, ʔy-, sɖ-, sm-, sʔ-, ʃq- are not listed in Li. Surprisingly, ʃt- and hm- listed in Prof. Li are not found in any of the data. The first member of the two consonant combinations is a voiceless stop p, t, k, q, ʔ, a voiceless fricative f, θ, s, ʃ, or the voiceless lateral ɬ: only m is voiced. The second member can be any consonant except ŋ, b, and ʃ. It is probably accidental to have no b or ʃ, since d and the other fricatives do appear as the second member in the combinations. Moreover, b and ʃ appear to follow another consonant in the word-medial position.

(2) In the word-final position, Prof. Li's statement still holds true, but the following four clusters can be added to his list: -ws, -wʃ, -wr and -yr.

As for combinations of three consonants in the word-medial position, the following can be added to those listed by Prof. Li: -qbɖ-, -qʃy-, -qry-, -mpl- and -nhy-.

3. Morphophonemic Alternations

3.1. Alternation of -m- ~ -um- (~-im- ~-am-)

The present tense is generally indicated by the infix -m- after the initial consonant of the verb base form which is usually

also the imperative form:[1]

(1) θ/m/a:nit[2] 'to cry' θa:nit 'cry!'

ʃ/m/i:naw 'to wash' ʃi:naw 'wash!'

However, -um- appears instead of -m- in:

(2) q/um/pit 'to cut'(a string)' qpi:t-i 'cut!'

cf. qpi:t-in 'is cut'

Note that -m- is historically derived from -um-.[3] The vowel of the infix has disappeared in most cases in Thao, but turned up to avoid #CCC, i.e. three consonants in a row word-initially.

Prof. Li (1956:28) noted the alternations of the same infix -m- ∼ -um- ∼ -im- ∼ -am- and illustrated with some examples. My recording of the same items often do not have the vowels of the infix. Cf. Prof. Li's transcription and mine (the slashes were added by me) below:

	Prof. Li's	Mine
(3)	t/um/á:la?	t/m/a:la 'to cut (wood)'
	r/úm/faɗ cf. má-rfaɗ 'to fly'	(same as the left) 'bird'
	t/úm/utu?	t/m/utu 'to suck milk'
(4)	t/im/í:tiʃ	t/m/i:tiʃ 'to wipe'
	k/im/in/ú:ɬaw	k/m/in/u:ɬaw 'to have scratched'
	k/im/in/a:ní:ɗa?	k/im/in/a:ni:ɗa, k/m/ina:ni:ɗa 'to have eaten'
(5)	t/am/áɬtaɬ	(missing) 'to split wood'

---

1. The imperative forms of most verbs are simply the base forms, e.g., θa:nit 'cry!' ʃi:naw 'wash!' pa?is?is 'rub!' A few verbs take the suffix -i or -y, e.g., qpi:t-i 'cut!' fari:w-i 'sell!' titi:ʃ-i 'wipe!' ta:na-y 'open the door!'
2. The slashes // indicate an infix.
3. Cf. Saisiyat -om- (see Li 1975).

The vowels of the infix are phonetically very weak and in fact optional, thus they are phonemically insignificant. Even if they should be represented in the phonemic transcription, they are mostly predictable and identical with the first vowel of the verb stem. Of the two exceptions, r/um/fad 'bird' is a noun and not a regular alternation in the verb paradigm; only t/um/a:la? 'to cut (wood)' as recorded by Prof. Li is unpredictable.

The phonetically weak vowels that are predictable can be inserted by an optional phonological rule:

$$\text{R5. } \emptyset \rightarrow \begin{bmatrix} +\text{syll} \\ \alpha\text{high} \\ \beta\text{back} \\ -\text{stress} \end{bmatrix} \Big/ [+\text{cons}] \_\_\_\_ [+\text{cons}] \begin{bmatrix} +\text{syll} \\ \alpha\text{high} \\ \beta\text{back} \end{bmatrix}$$

Let it be noted that the optional vowel addition rule above does not agree with the historical process: the loss of the infix vowel u.

3.2. Alternation of h ~ Ø

Observe the examples below:

(1) ta:la 'cut (wood)!' tala:h-an 'to be cut'

pa:qu 'eat/drink with!' ?a-paqu:h-an 'will be eaten/drunk with'

Note that the underlying forms of the verbs above end with h, which is deleted in the word-final position, but preserved when followed by a suffix. The final h deletion rule can be stated as below:

(2) h --> Ø/____#

(2) can be restated by features:

$$\text{R6. } \begin{bmatrix} +\text{cons} \\ -\text{ant} \\ -\text{cor} \\ +\text{cont} \end{bmatrix} \rightarrow \emptyset \,/ \_\_\_\_ \#$$

Since h never occurs word-finally in the language (see §2.5), this deletion rule will not affect any lexical item inappropriately.

3.3. Alternation of -in ~ -an ~ -n

The passive form of the verb is indicated by the suffix -in, -an or -n, as in Column II below:

	I. Active	II. Passive
(1)	k/m/a:n 'to eat'	ka:n-in 'to be eaten'
	θ/m/a:nup 'to bury'	θanu:p-in 'to be buried'
	makʃnaw 'to breathe'	pakʃnaw-in 'to be breathed(?)'
	fa:riw 'to buy/sell'	fari:w-in 'to be bought/sold'
	makɬamɬam 'to chew'	pakɬamɬam-in 'to be chewed'
	qɬi:t 'cut!'	qɬi:t-in 'to be cut'
	k/m/utawfaɬ 'to fight'	kutawfa:ɬ-in 'to be fought'
(2)	ta:la 'cut (wood)!'	tala:h-an 'to be cut'
	pa:qu 'to eat/drink with!'	ʔa-paqu:h-an 'will be eaten/ drunk with'
(3)	ka:θu 'to bring'	ka:θu-n 'to be brought'
	minfa:ri 'to blow'	ʔinfa:ri-n 'to be blown'
	paʃna:ra 'to burn'	paʃna:ra-n 'to be burned'

The above data indicate that verbs that end in a consonant including a semivowel take the passive suffix -in, verbs whose underlying forms end in h take -an (the vowel i is "pulled down" to a by the glottal fricative h), and verbs that end in a vowel take -n. If -in is treated as the underlying form of the passive suffix, then -an and -n can be derived by Rules (4) and (5) respectively:

(4) i --→ a/h+___n#

(5) i --→ Ø/V+___n#

The conditions in (4) and (5) cannot be simplified. For example, the morpheme boundary marker in (4) cannot be left out because not all i's after h or before n change to a, e.g., banhi:lan 'bed bug,' ba?hi 'soul,' ?inay 'here.' The n# in (5) cannot be left out because not all i's after a vowel followed by a morpheme boundary is omissible, e.g., ?ata qawriwa-i 'don't throw it away!'

In terms of features, (4) and (5) can be restated as R7 and R8 respectively:

R7. $\begin{bmatrix}+syll\\-back\end{bmatrix}$ --→ $[-high]$ / $\begin{bmatrix}+cons\\-cor\\-ant\\+cont\end{bmatrix}$ + ____ $\begin{bmatrix}+nas\\+cor\end{bmatrix}$ #

R8. $\begin{bmatrix}+syll\\-back\end{bmatrix}$ --→ Ø / $[+syll]$ + ____ $\begin{bmatrix}+nas\\+cor\end{bmatrix}$ #

The [+syll] before the + in R8 excludes a semivowel. For example, the form fa:riw does not apply, so we get fari:w-in, not *fa:riwn.

3.4. Alternation of -u:wan ~ -wan

Cf. (1) and (2) below:

(1) ?unay-wan 'please come'
rasirasi:-wan 'please bite'
tanwara:-wan 'please choose'
tma:da:-wan 'please listen'

(2) kan-u:wan 'please eat'
qilqil-u:wan 'please bite'
karkar-u:wan 'please chew'
?upruq-u:wan 'please come down'

The data above show that the suffixes -wan and -u:wan 'please, polite request' are phonologically conditioned; -wan

after a vowel or semivowel, and -u:wan after a consonant.[1] Such an alternation is to get the "preferred" canonical form CVCVC#. The problem is to determine which is the base. There are two possible solutions:

(3) $\emptyset \dashrightarrow u / C+\_\_\_\_$

(4) $u \dashrightarrow \emptyset / \begin{bmatrix} V \\ Y \end{bmatrix} +\_\_\_\_$ (Y indicates a semivowel).

It is simpler to delete u rather than insert it. The underlying form of the suffix is /-uwan/ (the vowel length and stress are late rules). (4) can be restated in terms of features:

R9. $\begin{bmatrix} +syll \\ +back \end{bmatrix} \dashrightarrow \emptyset / [-cons] +\_\_\_\_$

The rule above assumes that -u:wan is the only suffix that begins with u.

3.5. Alternations of i ~ y and u ~ w

The imperative is, as mentioned in §3.1, usually indicated by the base form of the verb, but sometimes by the base form plus the suffix -i if it ends in a consonant, or plus -y if it ends in a vowel (Li 1956, §2.401):[2]

(1)	I. Base	II. Imperative
(a)	ta:dam 'to taste'	tada:m-i 'taste!'
	ʃu:dap 'to touch'	ʃuda:p-i 'touch!'
	quθquθ 'to tie'	quθqu:θ-i 'tie!'

---

1. Prof. Li (1956, §2.405) gave a third suffix -i:wan with the same function, e.g., titiʃi:wan 'please wipe,' cf. t/im/i:tiʃ 'to wipe.' I believe that -i(:) is the imperative marker for a few verbs and a separate morpheme from -u:wan and -wan, i.e titiʃ-i:-wan, ʃaʃaw-i:-wan 'please chase it away.' See Note 1 in §3.1.

2. Many of the observations of the alternations and examples in this and the following sections are based on Li 1956, §2.4.

(b) ta:na 'to open the door' ta:na-y 'open the door!'

The i ~ y alternation also occurs in ʔi-náy 'here,' ʔa-y-náy 'will be here.'

The emphatic second person (?) is indicated by the suffix -u if the verb ends in a consonant, and by -w if it ends in a vowel (Li 1956, §2.402):

(2)	I. Imperative	II. Emphatic 2nd Person (?)
(a)	θa:nit 'cry!'	θanit-úʔ 'you cry!'
	ʃaqiʃ-i 'sew!'	ʃaqiʃ-úʔ 'you sew!'
(b)	ʔa:ra 'take!'	ʔara-ẃʔ 'you take!'

The u ~ w alternation is also manifested in the stems taking prefixes ending in a consonant or a vowel:

(3)	m-uʃu:qiʃ 'to turn'	ta-wʃu:qiʃ 'to return'
	m-unay 'to come'	ʃa-wna-nay 'to arrive'

The above data indicate that high vowels i and u become semivowels y and w respectively after a vowel. Although there is no example in the above data for a verb stem ending in a high vowel, it can be assumed that the same alternations would also hold true. In fact, the i ~ y alternation also takes place after a high vowel; see §3.6, (1)(b) below. Cf. the similar alternations in Rukai; see Li 1973:42-43.

These alternations in Thao can be stated by the rules below:

(4)(a) i --> y/V+____

(b) u --> w/V+____

The morpheme boundary marker in the rules cannot be left out because Thao allows vowel sequence, e.g., wiaqdas 'leech,' malalia 'to run.' In terms of features (4) can be restated in one rule:

R10. [+high] --> [-syll]/[+syll]+____

3.6. Vowel Deletion

The verbal suffixes -i:ɗa ~ -yɗa ~ -iɗa ~ -ɗa all indicate the completive aspect (or 'emphatic,' as stated in Li 1956, §2.408): -i:ɗa after a consonant, -yɗa after a back vowel, -iɗa after vowel i, and -ɗa after y:

(1)

	I	II
(a)	ma:ɬus 'to sleep'	m/in/aɬus-i:ɗa 'have slept'
	θanu:pin 'be buried'	θanupin-i:ɗa 'have been buried'
(b)	mi:lu 'to bathe'	m/in/i:lu-yɗa 'have bathed'
	ʔiŋqa 'don't'	ʔiŋqa-yɗa 'don't need!'
	piti:a 'to cook'	p/in/itia-yɗa 'have cooked'
(c)	miɬi:ɬi 'to stand'	m/in/iɬi:ɬi-iɗa 'have stood up'
(d)	ma:θay 'to die'	m/in/a:θay-ɗa 'have died'
	ʃ/m/awnanáy 'to arrive'	ʃ/m/in/awnanay-ɗa 'have arrived'

Ignoring the vowel length, which generally coincides with the stressed vowel and can be done by a late lengthening rule, the alternation between i and y was discussed in the previous section, §3.5. (1)(c) manifests an identical vowel deletion: i + i → i; the penultimate syllable is lengthened because of stress (R3). The examples in (1)(d) show the loss of the high front vowel after its counterpart high front semivowel y. Both (1)(c) and (d) can be captured by the rule:

R11. $\begin{bmatrix}-\text{cons}\\ \alpha\text{high}\\ \beta\text{back}\end{bmatrix} \dashrightarrow \emptyset \Big/ \begin{bmatrix}-\text{cons}\\ \alpha\text{high}\\ \beta\text{back}\end{bmatrix} +\_\_\_\_$

The above rule assumes that a + a --> a, u + u --> u, w + u --> u, etc. will also take place. The deletion of the low vowel is attested in the alternation of the suffixes -ak ~ -k, which indicate 'the first person, myself,' as in the following examples:

(2)	I	II
(a)	m-riqad 'to see'	ʔa-riqad-ak 'I shall see'
(b)	m-iqi:ɬa 'to drink'	ʔa-yqi:ɬa-k 'I shall drink'
	m-a:pa 'to carry'	ʔa-pa:pa-k 'I shall carry'

The above data show that -ak occurs after a verb ending in a consonant and -k after a verb ending in the low vowel a. This is clearly a case of a + a --> a.[1]

Although no example of identical high back vowel deletion is attested (cf. -u:wan ~ -wan in §3.4), it can be so assumed. It would be very unnatural if that was not the case.

Identical vowel sequences aa, ii and uu do not occur in Thao, but there are sequences of uw, iy, etc. These will be affected if the morpheme boundary marker in R10 is left out.

1. Prof. Li (1956, §2.409) gave a third suffix -ik, e.g., ʔa-ka:n-ik 'I shall eat.' The i in -ik is irregular and perhaps assimilated to the preceding vowel which has been lost: PAN *kaʔən 'eat' +*aku 'I' > *kaʔən-əku > *kaʔin-ik > ka:n-ik 'I eat.'

Bibliography

Abe, Akiyoshi. 1930. 蕃語研究 [Study of Languages of Native Tribes]. Taipei: Bango Kenkyûkai.

Alvarez, Fr. José María (O.P.) 1915-24. Descripción Geográfica de la Isla de Formosa. Madrid: Real Sociedad Geográfica.

Bullock, T.L. 1874-5. "Formosan Dialects and Their Connection with the Malay," China Review 3:38-46. Hong Kong.

Chen, Chi-lu, Yih-yüan Li and Mei-chün Tang. 1955. "Preliminary Report of Ethnological Investigation in the Thao of Jih Yüeh T'an (Sun-Moon Lake), Taiwan," Bulletin of the Department of Archaeology and Anthropology, National Taiwan University 6:26-33.

Collingwood, Dr. 1868. "Visit to the Kibalan Village of Sau-o Bay, Northeast Coast of Formosa," Transactions of the Ethnological Society N.S. 6:135-43, 362-63 (vocabulary). London.

Dyen, Isidore. 1963. "The position of the Malayopolynesian Languages of Formosa," Asian Perspectives 7(1/2):261-71.

---. 1971. "The Austronesian Languages of Formosa," Current Trends in Linguistics 8:168-99, ed. by Thomas Sebeok.

Li, Fang Kuei, Chi-lu Chen and Mei-chün Tang. 1956. "Notes on the Thao Language," Bulletin of the Department of Archaeology and Anthropology, National Taiwan University 7:23-51.

Li, Paul Jen-kuei. 1973. Rukai Structure. Institute of History and Philology, Academia Sinica Special Publications No.64. Taipei

---. 1974. "Alternations between Semiconsonants and Fricatives or Liquids." Paper presented at the First International Conference on Comparative Austronesian Linguistics, Honolulu. To

appear in Oceanic Linguistics 13.

---. 1975a. Saisiyat Structure and Texts. Unpublished manuscripts.

---. 1975b. Thao Texts. Unpublished manuscripts.

Ogawa, Naoyoshi and Erin Asai. 1935. 原語による台灣高砂族傳說集 [Myths and Traditions of the Formosan Native Tribes]. Taihoku: Taihoku Teikoku Daigaku Gengo-gaku Kenkyû-shitsu.

Steere, J. B. 1874. "[The Aborigines of] Formosa," Journal of the American Geographical Society of New York 6:302-34.

Tsuchida, Shigeru. 1975. Reconstruction of Proto-Tsouic Phonology. Unpublished PhD dissertation at Yale University. New Haven.

Tung, T'ung-ho. 1964. A Descriptive Study of the Tsou Language, Formosa. Taipei: Institute of History and Philology, Academia Sinica Special Publications No.48.

「邵語音韻」中文摘要

邵族號稱臺灣高山族十族之一，但目前邵語已接近消滅的邊緣，而有關邵語的資料跟研究極少。連厚達八百多頁的原語による臺灣高砂族傳說集都沒有邵語的資料。舊文獻雖列有少數邵語的單字，但記音並不準確。

有關邵語的報告最好的是李方桂師等人所寫的「邵語記略」，對邵語的語音及構詞有簡略的描述。雖然他們在日月潭只收集了五個半天的資料，但其記音非常精確，對結構的觀察也很洞澈。報告中所列的約八百個單字已成爲南島語言比較研究的主要邵語資料來源。可惜那些資料還不够。因此亟需再去做田野調查，收集更多的資料。

作者在一九七五年六、七月間在日月潭德化社工作了三週，收集了二十六則語料（打字及譯文共75頁），單字也收了約一千五百個字，此外還收了一些單句。這篇報告只是有關邵語調查的一部份研究心得而已。

本文的目的之一是把邵語音韻研究較詳盡的結果提供給更多的讀者。「邵語記略」一文的詞彙部份雖附有英文翻譯，但報告的全文是用中文寫的，因此不懂中文的人只能使用其詞彙資料而已。本文對邵語音韻系統有較詳盡的探討與說明。

邵語共有二十一個輔音 p, t, k, q, ʔ, b, d, m, n, ŋ, f, θ, ð̸ (ð), s, ʃ, h, l, ɬ, r, w, y, 只有三個元音 a, i, u。輔音比「邵語記略」所列的多了一個 s，因 s 與 θ 成對比。元音却去掉了 e，因 e 只是 i 元音的變音。這是掌握了較豐富的資料後所得的結果。從平面的或歷史的觀點去探討，這些修正似乎是合理的。此外，本文所採用的產生音韻學理論跟十九年前「邵語記略」所採用的結構音韻學理論在方法上也不盡相同。譬如，詞音位轉換 (morphophonemic alternations) 的現象原來在構詞部份需要個別列舉的，現在都在音韻部份同時處理，只要幾條規律就可以一起獲得解決了。然而，本文的寫作得力於「邵語記略」的啓發非常之多，這是作者要特別誌謝的。

本文分下列幾節：一、概說，二、音韻：1. 輔音元音表，2. 語音描述（包括辨音徵性表），3. 重音，4. 長音，5. 語音分佈，三、詞音位轉換：1. 插詞（中加成分）

-m-～-um- (～-im-～-am-) 的轉換，2. 字尾 h～ϕ 的轉換，3. 詞尾 -in～-an～-n 的轉換，4. 詞尾 -u:wan～-wan 的轉換，5. 高元音與半音的轉換 i～y, u～w, 6. 相同元音串的刪減。

李 壬 癸