

# Thao

*Paul Jen-kuei Li and Elizabeth Zeitoun*

58.1	Introduction	2
58.1.1	<i>Geographical Location and Population</i>	2
58.1.2	<i>Position among the Formosan Languages</i>	2
58.1.3	<i>Documentation</i>	3
58.2	Phonology	4
58.2.1	<i>Phonemic Inventory and Orthographic System</i>	4
58.2.2	<i>Distribution</i>	6
58.2.3	<i>Syllable Structure and Stress</i>	8
58.3	Morphology	9
58.3.1	<i>Morphological Units</i>	9
58.3.2	<i>Morphological Processes</i>	9
58.4	Word Classes	11
58.5	Syntax	12
58.5.1	<i>An Overview of Basic Clause Structure</i>	12
58.5.2	<i>Noun Phrase Structure</i>	14
58.5.3	<i>Personal Pronouns</i>	17
58.5.4	<i>Voice, Aspect, and Mood System</i>	20
58.5.5	<i>Negation</i>	24
58.5.6	<i>Interrogatives</i>	25
58.5.7	<i>Transcategorical Operations</i>	27
58.5.8	<i>Valency-Adjusting Operations</i>	28
58.6	Complex Sentences	30
58.6.1	<i>Serial Verb Constructions</i>	30
58.6.2	<i>Complementation</i>	31
58.6.3	<i>Adverbial Clauses</i>	31
58.6.4	<i>Coordination</i>	32
	Acknowledgments	33
	References	33

## 58.1 Introduction

The present chapter describes Thao phonology (phoneme inventory, orthography, and sound patterns), word formation, word classes, and major syntactic structures. Thao displays some interesting linguistic features that are either not found or only rarely found in the other Formosan languages. It has borrowed extensively from Bunun, and as a consequence, five consonants /<sup>2</sup>b <sup>2</sup>d l ʔ h/ have now acquired full phonemic status in Thao (Li 2013; see § 58.2). Phonologically, Thao has the largest number of fricatives of any Formosan language and a great number of consonant clusters, which make accurate transcriptions challenging. Morphologically, it has over 200 affixes (Blust 2003a, pp. 91–198) and many possible combinations thereof. Syntactically, it has no circumstantial voice, and there are only two bound pronominal forms.

### 58.1.1 *Geographical Location and Population*

Thao is spoken at Sun Moon Lake in central Taiwan. It was officially recognized as the tenth indigenous ethnic group of Taiwan on 22 September 2001, just two years after the devastating magnitude 7.3 Jiji earthquake on 21 September 1999, in Nantou County. As of August 2020, Thao had an ethnic population of fewer than 850 people (Council of Indigenous Peoples 2020), most of whom reside in the De-hua Village, renamed Ita Thao (Barawbaw in Thao) in Yuchi Township, Nantou County. Blust (2003a, p. 4) mentions rightfully that “[t]he present situation of the Thao can be described as one of terminal assimilation. Tourists are told about the ‘Thao village’ at Sun Moon Lake, but no ethnically distinct village exists. Rather, the Thao are a culturally and linguistically vanishing minority in a Taiwanese-speaking village where their aboriginal heritage is liberally exploited for commercial purposes through the conspicuous marketing of gaudy tourist memorabilia.” Thao is a critically endangered language with only a few speakers left today, none of whom is really fluent, despite attempts at revitalization.

### 58.1.2 *Position among the Formosan Languages*

Until the mid-1990s, the position of Thao among the Formosan languages remained unclear. Ferrell (1969, p. 25) proposed dividing the Formosan languages into three groups, Atayalic (Atayal and Seediq), Tsouic (Tsou, Kanakanavu and Saaroa) and Paiwanic (all the remaining Formosan languages) and classified Thao as part of the Paiwanic group. More specifically, Thao was included in Paiwanic I, along with Rukai, Pazeh, and Saisiyat.

Blust (1996) was the first to clearly identify Thao, based on lexical evidence, as genetically closer to the sinicized languages known as the languages of the

Western Plains (WP) (Taokas, Babuza, Hoanya, and Papora). As he pointed out, Thao shares “certain lexical innovations exclusively with one or more WP languages” (p. 279), for instance PWP \*maka-Sepat ‘eight’: Taokas *maka-apat*, Babuza *maa-spat*, Thao *maka-shpa~shpat* ‘eight’ (as opposed to PAN \*walu ‘eight’).<sup>1</sup>

Li (2001) reassessed the reconstructions proposed by Blust and concluded that Thao and the WP languages exclusively share two phonological innovations, PAN \*ŋ > n and \*s > t (except word-finally in Thao). Thao, however does not share the following innovations with the WP languages: PAN \*k > Ø and \*y > Ø.

### 58.1.3 Documentation

In spite of the proliferation of studies on Thao since the mid-1990s, it is not as well studied as many other Formosan languages, and no in-depth reference grammar on the language has appeared to date.

The first important linguistic work on Thao was by F. K. Li et al. (1956), which contains a brief description of Thao phonology, morphology, and grammatical particles, followed by an approximately 800-item word list and two short texts. P. Li (1976) amended the phonological system of Thao by demonstrating, among others, the phonemic contrast between /s/ and /θ/. Blust treats *q* as pharyngeal instead of uvular; he believes that *b* and *d* should be regarded as preglottalized and not implosive stops; he describes all of the dentals except for *d* as postdentals; and he analyzes *r* as a flap instead of a trill. He deals with a number of phonological problems, which include the merger of *q* vs. *k*; the confusion between *z*, *l*, and *r*; the number of vowels; stress shift; vowel syncope; vowel lowering; allomorphy of <um> and <in>; haplology; and sibilant assimilation. Chang (1998) was the first to deal with reduplication in Thao and demonstrated that this language is very rich in terms of patterns of reduplication. This topic was taken up again by Lu (2003) and Lee (2007). Among the various studies by Blust (1996, 1998a, 1998b, 2001, 2003a, 2000b), his *Thao dictionary* (2003a) is by far the most comprehensive and important. The 240-page introduction provides a very detailed account of Thao phonology, both synchronic and diachronic, as well as morphology and grammatical systems, plus five texts collected from three different speakers. The main body of the dictionary provides unusually rich lexical data and examples over nearly 800 pages. Abe's dictionary (2008) is much smaller in scope, listing only 3,597 lexical forms,

1 Blust (1999) added another language, Pazeh, into the Western Plains group, but this is still a controversial issue.

some of which are illustrated with examples, over some 200 pages. An online Thao-Chinese dictionary is available, edited by Shi-Lang Jean and sponsored by the Council of Indigenous Peoples, and a trilingual Thao-Chinese-English dictionary by Li is forthcoming. Four MA theses have been written on Thao: Youmehim Chen (2000) on negation; Weng (2000) on tense, aspect, and mood; Tseng (2008) on morphophonemic alternations; and Yu-chuan D. Chen (2014) on nominalization. There is one PhD dissertation, Wang (2004), which is by far the most comprehensive study of Thao syntax to date, with an account of basic clause structure, transitivity, and ergativity. A few papers on disparate phonological and morphosyntactic topics can also be found (see Li 2013, 2014, 2016, Niida 2007, and Tsuchida 1989), as can two sketch grammars written in Chinese (Huang 2000, Jean 2018) and a study on Thao word classes (Zeitoun 2022).

Vocabulary, songs, and texts have been collected, which include Tsuchida (1989), Abe, Nagashima & Niida (2007, 2008), Li & Wu (2003), and Li (2011), the latter of which provides the largest number of texts and songs (42 texts and 19 songs) with interlinear glosses.

## 58.2 Phonology

This section gives a brief introduction to Thao phonology, with an overview of its phoneme inventory and orthographic system (§ 58.2.1), the distribution of vowels and consonants (§ 58.2.2), and syllable structure and stress (§ 58.2.3). Blust (1996, 2003a) offers a more detailed description, and the reader is referred to his work for details on morphophonemic alternations and phonological rules.

### 58.2.1 *Phonemic Inventory and Orthographic System*

Thao has 18 consonants /p t k q ʔ b ʔd m n f θ ð s ʃ h l ʎ r/, 2 semivowels /w j/, and 3 vowels /i u a/, as shown in Tables 58.1 and 58.2. The orthographic symbols are indicated in italics, followed by the IPA symbols placed in slashes to their right, based on the standardized orthography promulgated by the Council of Indigenous Peoples (CIP) and the Ministry of Education (MOE) (2005).

Among these, the five consonants /ʔ b ʔd l ʎ h/ originated from extensive lexical borrowing from Bunun, but are now distinct phonemes, as shown by the following near-minimal pairs: *bunaz* /ʔbunað/ ‘sand’ vs. *furaz* /furað/ ‘moon’, *daul* /ʔdaul/ ‘Asian snakehead (*Channa asiatica* Linn.)’ vs. *taun* /taun/ ‘house’, *lamin* /lamin/ ‘fortunately’ vs. *lhamith* /ʎamiθ/ ‘root’, *itia* /itia/ ‘to have’ vs. *pit’ia*

TABLE 58.1 Thao consonants

		(Bi)labial	Interdental	Dental	Palatal	Velar	Uvular	Glottal
Stops	VL	<i>p</i> /p/		<i>t</i> /t/		<i>k</i> /k/	<i>q</i> /q/	<i>ʔ</i> /ʔ/
	VD	<i>b</i> /ʔb/		<i>d</i> /ʔd/				
Nasal		<i>m</i> /m/		<i>n</i> /n/				
Fricative	VL	<i>f</i> /ɸ/	<i>th</i> /θ/	<i>s</i> /s/	<i>sh</i> /ʃ/			<i>h</i> /h/
	VD		<i>z</i> /ð/					
Lateral	VL			<i>lh</i> /ɬ/				
	VD			<i>l</i> /l/				
Trill				<i>r</i> /r/				
Approximant		<i>w</i> /w/				<i>y</i> /j/		

TABLE 58.2 Thao vowels

	Front	Central	Back
High	<i>i</i> /i/		<i>u</i> /u/
Mid			
Low		<i>a</i> /a/	

/pitʔia/ ‘to cook’, *isáy* /isay/ ‘there’ vs. *isaháy* /isahaj/ ‘to stay there’. The voiceless stops are unaspirated, while *b* and *d* as preglottalized voiced stops [ʔb ʔd] sound like implosives, as in Bunun and Tsou (Li 1976, p. 223). Unlike function words (see *sa* ‘CN’, =*a* ~ =*wa* ‘LNK’, *na* ‘NOM’, *tu* ‘LNK’, *ti* ‘PN’), every content word in Thao begins and ends with a phonetic glottal stop if no other consonant is present, e.g., *uka* [ʔukaʔ] ‘not have’. Since the glottal stop is usually phonetically predictable, it is not represented in such environments, with the only exception being *qriuʔ* /qriuʔ/ ‘to steal’, according to Blust (2003a, p. 23). As shown above, however, the glottal stop is phonemic in word-middle position, even if only found in contrastive loanwords.

There are only two nasal phonemes /m n/. The velar nasal *ng* /ŋ/ is extremely rare, e.g., *pishtingting* /piftiŋtiŋ/ ‘to be upset’, and usually occurs in names and onomatopoeic words (Blust 2003a, p. 24). It is also usually phonetically realized as a velar nasal when immediately followed by a velar or uvular stop, e.g., *pangka* /paŋka/ ‘table’, *pangqa* /paŋqa/ ‘to rest’, and can be treated as a loan phoneme from Bunun.

Thao exhibits an unusually large number of fricatives. The fricative *f* /f/ is treated as a bilabial fricative by Blust (2003a, p. 31). The two consonants *th* /θ/ and *z* /ð/ are interdental fricatives. There is a voiceless dental lateral fricative *lh* /ɬ/, which contrasts with a voiced lateral approximant *l* /l/. Due to the large number of voiceless sibilants in Thao, there is a strong tendency for one to assimilate to the other, e.g., *lhizashan* /ɬiðʰaʃan/ ~ *lhilhashan* /ɬiɬaʃan/ ‘pheasant’, *falhuz* /faɬuð/ ~ *falhuɬ* /faɬuɬ/ ‘green pigeon’, *lhmaushin* /ɬmauʃin/ ~ *lhmaulhin* /ɬmauʃin/ ‘to swing’ (Blust 2003a, pp. 69–70). Such variation is also found in the transcriptions of different fieldworkers. For instance, the form *shupilh* /ʃupiɬ/ ‘to count’ was recorded by P. Li<sup>2</sup>, while Blust (2003a, p. 350) has *thupilh* /θupiɬ/, in which the initial /ʃ/ has been replaced by another fricative, /θ/. The semivowel *w* /w/ is realized phonetically as a voiced bilabial fricative [β] pre-vocally and as a semivowel [w] post-vocally.

The three vowels *i*, *u*, and *a* /i u a/ have their expected values, but the high vowels *i* /i/ and *u* /u/ are lowered when preceding or following certain consonants. More specifically, /i/ is pronounced as a mid-vowel [e] when adjacent to /r h ʔ/, as in *lhmirik* /ɬmirik/ [ɬmérek] ‘to pierce’, *kunʔishʔishir* /kunʔiʃiʔiʃir/ [kunʔiʃiʔiʃer] ‘to be incontinent’, *mashʔia* /maʃʔia/ [maʃʔeaʔ] ‘to have a priest-shaman perform a ritual’; /u/ is lowered to [o] when adjacent to /q r ŋ/, e.g., *quyash* /quyaʃ/ [qoyaʃ] ‘song’, *roza* /ruða/ [roðaʔ] ‘boat’, *ilhungqu* /iɬuŋqu/ [iɬoŋqoʔ] ‘to sit’. In addition, /i/ and /u/ are realized phonetically as mid vowels [e] and [o], respectively, with a transitional schwa when adjacent to /q/.

### 58.2.2 Distribution

Without going into too much detail, except for *b* /<sup>2</sup>b/, *d* /<sup>2</sup>d/, *h* /h/, and ' /ʔ/, all consonants may occur in word-initial, word-medial, and word-final positions. The consonants *b* /<sup>2</sup>b/ and *d* /<sup>2</sup>d/ only occur in syllable-initial position, while *h* /h/ does not occur in word-final position, though it resurfaces when a word is suffixed; compare *bizu* /<sup>2</sup>biðu/ ‘beard’ and *tan-bizuh-an* /tan<sup>2</sup>biðuhan/ ‘bearded’, *tala* /tala/ ‘to chop (wood)’ and *talah-an* /talahan/ ‘what is chopped’. The glottal stop, as shown earlier, is only found phonemically in word-medial position. The restriction of the occurrence of these consonants may have to do with their appearance mainly in loanwords from Bunun (Li 2013). There are no restrictions on the distribution of the vowels.

2 It is regularly derived from PAN \*SupəR ‘to count’; cf. Seediq *s<m>epug* /sməpug/, Tsou *s<m>upru* /smupru/, Kanakanavu *s<um>a-supuru* /sumasupuru/, Bunun *ma-sipul* /masipul/, Paiwan *s<em>upu* /səmpu/, Saisiyat *shepeL* /ʃəpər/. Blust & Trussel (2022) reconstruct this form as PAN \*SipuR, which could be revised as PAN \*SupeR, based on the reflexes in the majority of Formosan languages. Alternatively, PAN \*SipuR and \*SupeR could be treated as doublets.

Thao, along with Tsou and Maga Rukai, is one of a few Formosan languages that have true consonant clusters, which occur in word-initial and word-medial positions, but never in word-final position. In word-initial position, there are combinations of (i) a stop (with the exception of *d* /ʔd/) and a stop, nasal, liquid, or fricative, e.g., *pt-*, *pn-*, *pl-*, *pr-*, *pʰ-*, *pf-*, *ph-*, *kp-*, *kt-*, <sup>2</sup>*bl-*, <sup>2</sup>*br-*, <sup>2</sup>*bs-*, <sup>2</sup>*bz-*, *tk-*, *tq-*, *tʰd-*, *tm-*, *tr-*, *th-*, *km-*, *kn-*, *kθ-*, *kð-*, *qp-*, *qt-*, *qʰb-*, *qʰd-*, *qm-*, *qn-*, *ql-*, *qr-*, *qf-*, *qθ-*, *qð-*, *qʰ-*, *qʃ-*; (ii) a fricative (with the exception of *z* /ð/) and a stop, nasal, or liquid, e.g., *θp-*, *θk-*, *θm-*, *fl-*, *fr-*, *ft-*, *fq-*, *sp-*, *sk-*, *sʰb-*, *sm-*, *sn-*, *ʃp-*, *ʃt-*, *ʃk-*, *ʃq-*, *ʃʰb-*, *ʃʰd-*, *ʃm-*, *ʃn-*, *ʃl-*, *ʃr-*; (iii) fricatives, e.g., *fð-*, *ʃð-*. Blust (2003a, p. 20) states that “perhaps nasals, liquids and glides may not appear as cluster onsets.” However, /*m*/ is found as a cluster onset followed by a liquid in a number of forms, e.g., *mrafið* ‘to fan’, *mlalas* ‘to peel with a knife’. A few (true) trilateral consonant clusters appear word-initially, cf. *#ʔθk-*, *#ʔqn-*, as in *lhthkiz* /ʔθkiz/ ‘at one time’, *lhqniz* /ʔqniz/ ‘to bear down’, and word-medially (of the type C.CCV), *-ŋqt-*, *-ŋqθ-*, *-ŋkm-*, *-ŋql-*, *-ŋqʰ-*, *-kʰbð-*, as in *angqtu* /aŋqtu/ ‘to contemplate’, *qungqthu* /quŋqθu/ ‘to eat delicacies’, *ingkmir* /iŋkmir/ ‘to roll into a ball’, *tangqlin* /taŋqlin/ ‘plant sp., *Gardenia jasminoides* Ellis’, *kungqlha* /kuŋqla/ ‘to cry out (AV)’, *pikbaw* /pikʰbðaw/ ‘to conquer’ (see also Blust 2003a, pp. 22–23). Only unlike consonants or vowels are permitted adjacent to one another. When a sequence of identical consonants or vowels occur across a morpheme boundary, it is usually simplified to a single consonant or vowel, e.g., *mash-shput* /maʃʃput/ > *mashput* /maʃput/ ‘to speak Taiwanese’, *ilhungquuan* /iʔuŋquuan/ > *ilhungquan* /iʔuŋquan/ ‘Sit down, please!’. A single nasal in intervocalic position is pronounced phonetically as a geminate, e.g., *thumay* /θumaj/ [θummaj] ‘bear’. A few lexical forms exhibit sporadic metathesis, e.g., *shupilh* /ʃupiʔ/ ~ *lhupish* /ʔupif/ ‘to count’, *sapuk* /sapuk/ ~ *sakup* /sakup/ ‘to catch’.

Tables 58.3 and 58.4 illustrate the distributions of the consonants and vowels in Thao.

TABLE 58.3 Distribution of Thao consonants

	Word-initial	Gloss	Intervocalic	Gloss	Word-final	Gloss
<b>p</b>	<i>pali</i>	‘wing’	<i>tapish</i>	‘to winnow’	<i>iup</i>	‘to blow’
<b>t</b>	<i>taina</i>	‘female’	<i>kitaz</i>	‘to bow’	<i>fufut</i>	‘flute’
<b>k</b>	<i>kuku</i>	‘fingernail’	<i>kukulay</i>	‘bugs’	<i>tmiktik</i>	‘chop meat (A
<b>q</b>	<i>qpit</i>	‘to press’	<i>mriqaz</i>	‘to see’	<i>paqsaq</i>	‘heel’
<b>’</b>	—	—	<i>mara’in</i>	‘big’	—	—
<b>b</b>	<i>buqtur</i>	‘neck, throat’	<i>bibi</i>	‘chin’	—	—

TABLE 58.3 Distribution of Thao consonants (*cont.*)

Word-initial	Gloss	Intervocalic	Gloss	Word-final	Gloss
<b>d</b>	<i>dudur</i>	'row'	<i>hudun</i>	'clothes'	—
<b>m</b>	<i>matha</i>	'eye'	<i>rima</i>	'hand'	<i>thazum</i>
<b>n</b>	<i>nazaq</i>	'pus'	<i>bunaz</i>	'sand'	<i>taun</i>
<b>f</b>	<i>fukish</i>	'hair'	<i>ifafaw</i>	'above'	<i>qaruf</i>
<b>s</b>	<i>suzay</i>	'reluctant'	<i>masasuga</i>	'quarrel'	<i>sukus</i>
<b>θ</b>	<i>thawa</i>	'to laugh'	<i>quthaz</i>	'rain'	<i>hukuth</i>
<b>z</b>	<i>zifu</i>	'nest'	<i>zazinis</i>	'honey'	<i>furaz</i>
<b>sh</b>	<i>shaish</i>	'take turn'	<i>tusha</i>	'two'	<i>funuf</i>
<b>h</b>	<i>harbuk</i>	'fog'	<i>humhum</i>	'twilight'	—
<b>l</b>	<i>lalay</i>	'cicada'	<i>lalalas</i>	'peeler'	<i>hulhul</i>
<b>lh</b>	<i>lhimaf</i>	'fat, lard'	<i>lhilhiq</i>	'to pull'	<i>futulh</i>
<b>r</b>	<i>rifi</i>	'liver'	<i>matarictic</i>	'blunt'	<i>fuar</i>
<b>w</b>	<i>waga</i>	'horn'	<i>kawi</i>	'wood'	<i>fariw</i>
<b>y</b>	<i>yanan</i>	'to have'	<i>ayaz</i>	'termite'	<i>usuhuy</i>

TABLE 58.4 Distribution of Thao vowels

Word-initial	Gloss	Word-medial	Gloss	Word-final	Gloss
<b>i</b>	<i>ina</i>	'mother'	<i>fizfiz</i>	'banana'	<i>tuali</i>
<b>u</b>	<i>utaq</i>	'to vomit'	<i>bunlhaz</i>	'flesh'	<i>kathu</i>
<b>a</b>	<i>afu</i>	'rice'	<i>kawash</i>	'year'	<i>taniza</i>

58.2.3 Syllable Structure and Stress

The most frequent syllable structures found in Thao content words include CV, CVC, and CCV. Other types of syllables are VC, V, CCVC, and, rarely, CCCV(C). Function words may be monosyllabic or disyllabic, e.g., =a ‘LNK’, *maqa* ‘because’. Content words are usually disyllabic or polysyllabic, e.g., *qusum* ‘to be dark, black’, *qatitira* ‘flea’. If a content word is monosyllabic, it is always bimoraic, i.e., the nucleic vowel is lengthened, e.g., *faq* [fa:q] ‘lung’, *qtut* [qtu:t] ‘fart’.

Stress is phonemic in Thao, although its functional load is not high. Blust (2003a, p. 35) gives the following contrastive pairs of examples: *hulus* /húlus/ ‘clothes’ vs. *tufush* /tufúʃ/ ‘sugarcane’, *kalhan* /káʔan/ ‘freshwater crab’ vs. *falhan*



/faʎán/ ‘ribs’, *tilkun* /tílkun/ ‘onion’ vs. *tikun* /tikún/ ‘kind of deep pot’. It usually falls on the penult, but appears on the final syllable in a few stems, such as *munáy* /munáy/ ‘come (AV)’, *a=makán* /amakán/ ‘(will) eat (AV)’. If stress is on the penultimate syllable in a stem, there is stress shift when it is suffixed (see Blust (2003a, pp. 35–41) for a detailed discussion of the stress-shift rules).

### 58.3 Morphology

As a synthetic agglutinative language, Thao is morphologically complex. In this section, we deal with morphological units (§58.3.1) and morphological processes (§58.3.2).

#### 58.3.1 Morphological Units

Relevant morphological units include roots, stems, affixes, clitics, and words. A root consists of a single morpheme, which can be either free (e.g., *matha* ‘eye’) or bound (e.g., *|kan|* ‘to eat’). A stem may consist of a free root by itself, e.g., *rauz* ‘swim’, or a (bound or free) root co-occurring with affixes or clitics, e.g., *|rawath|* ~ *min-rawath* ‘do something habitually, as an occupation’, *a=ma-rauz* ‘will swim’. A word may be a root or a stem with or without affixes. Affixes consist mostly of lexical affixes, which convey various meanings changing or reinforcing the connotation of the base to which they attach; there are fewer than a dozen grammatical affixes, which encode voice, mood, aspect, etc. Clitics, which only carry a grammatical function, are far less numerous than affixes and include proclitics, e.g., *a=* ‘IRR’, and enclitics, e.g., *=uan* ~ *=wan*<sup>3</sup> ‘still, please (in imperative clauses)’, and *=iza* ~ *=yza* ~ *=za*<sup>4</sup> ‘already’. Simple words can be monosyllabic (e.g., *qun* ‘mushroom’), disyllabic (e.g., *taun* ‘house’), or polysyllabic (e.g., *qarithuy* ‘egg’). Examples of complex words include *mak-lhu-lhun* ‘keep on blowing the nose (AV)’ (< *lhun* ‘wet nasal mucus’) and *p<in>an-saqazi* ‘afternoon nap’, consisting of the root *saqazi* ‘noon’, the verbalizer *pan-* (< *pan-saqazi* ‘to take an afternoon nap’), and the perfective nominalizer <*in*>.

#### 58.3.2 Morphological Processes

Thao exhibits two productive morphological processes, affixation and reduplication.

3 These allomorphs are phonologically conditioned, with *=uan* occurring after a consonant and *=wan* elsewhere.

4 These bound forms are phonologically conditioned: *=iza* occurs after a consonant, *=za* after *i* or *y*, and *=yza* elsewhere.

Nearly 200 prefixes and a few infixes, suffixes, and circumfixes were recorded by Blust (2003a, pp. 91–202), who provides a detailed account. When the two common infixes,  $\langle um \rangle$  and  $\langle in \rangle$ , or their variants co-occur, they appear in that order, e.g.,  $k\langle m \rangle\langle in \rangle an$  ‘have eaten’. In addition to these two productive infixes, Thao has three fossilized infixes,  $\langle ar \rangle$ ,  $\langle al \rangle$ , and  $\langle az \rangle$ , e.g.,  $q\langle ar \rangle afqaf$  ‘house’,  $sh\langle al \rangle inshin$  ‘bell’,  $k\langle az \rangle ingkin$  ‘earring’ (see Li & Tsuchida 2009, pp. 346–347 for more examples). Three affixes,  $\langle in \rangle/in-$  ‘PFV.UVP/PFV.PAT.NMLZ’,  $-in$  ‘UVP/PAT.NMLZ’, and  $-an$  ‘UVL/LOC.NMLZ’, have portmanteau functions as voice markers and nominalizers. They are used productively to derive patientive and locative nominals, e.g.,  $p\langle in \rangle arbu$  ‘which was baked’ <  $parbu$  ‘to bake’,  $in-apa$  ‘what was carried’ <  $apa$  ‘carry’,  $r\langle in \rangle iqaz-an$  ‘what was seen’ <  $riqaz$  ‘to see’,  $kan-in$  ‘food’ <  $|kan|$  ‘to eat’,  $kalhus-an$  ‘bed’ <  $kalhus$  ‘to sleep’ (see §58.5.6). Instrumental nouns, however, are derived through reduplication, which may occur independently or in co-occurrence with affixation, and has been discussed in great detail by Chang (1998), Lu (2003), and Lee (2007, pp. 337–340), who also outline the implications for phonological theory. Besides the numerous instances of lexicalized reduplication (e.g.,  $karkar$  ‘to chew’,  $bulbul$  ‘dust’), there are three productive patterns, full reduplication, Ca-reduplication, and rightward reduplication; two other patterns, CV-reduplication and triplication, are much less commonly used. Full reduplication consists of the reduplication of two syllables at most with the exclusion of the coda.<sup>5</sup> It marks repetitive or iterative aspect in dynamic verbs or derived nominals, e.g.,  $kau-kaush$  ‘scoop repeatedly’ <  $kaush$  ‘to scoop’,  $mu-buha~buhat$  ‘keep cultivating (AV)’ <  $buhat$  ‘field, to cultivate’,  $thupi~thupish-an$  ‘school’ (lit. ‘place where one often counts’) <  $thupish$  ‘to count’; intensity or collectivity in stative verbs, e.g.,  $ma-qita~qitan$  ‘very pretty’ <  $ma-qitan$  ‘good, pretty’; collective, e.g.,  $ma-puzi~puzi$  ‘all white’ <  $ma-puzi$  ‘white’; and plurality in nouns, e.g.,  $numa~numa$  ‘all sorts of things’ <  $numa$  ‘thing’. Ca-reduplication refers to the reduplication of the first consonant plus the vowel /a/ or, if there is no onset, the insertion of the vowel /a/. It applies to verbal, nominal, and numeral bases. The Ca-reduplication of a verb root primarily indicates an instrument, e.g.,  $fa-finshiq$  ‘seeds’ <  $finshiq$  ‘to sow’,  $ta~tiuz$  ‘a comb’ <  $tiuz$  ‘to comb’,  $lha~lhiklhik$  ‘a saw’ <  $lhiklhik$  ‘to saw’. However, Ca-reduplication also encodes repetitive or continuous aspect, e.g.,  $k\langle m \rangle a~kiskis$  ‘keep on pressing down’ <  $k\langle m \rangle iskis$  ‘press down (AV)’,  $pi-ta~t’ia-an$  ‘cooking place, kitchen’ <  $pit’ia$  ‘to cook’. In nouns, it conveys the meaning of ‘odor/smell of X’, as in  $tu-sha~shibun$  ‘odor of sweat’

5 One exception is *kudush~kudush* ‘nervous’, which might be a loan from Bunun (Blust 2003a, p. 489).

< *shibun* 'sweat' (see Lee 2010, 2021), and in numerals, *Ca*-reduplication is used to count people, e.g., *ta~turu* 'three people' < *turu* 'three'. It may also derive verbs with a human referent, e.g., *pu-lha~lhalhuzu* 'person specialized in setting fish traps' < *pu-lhalhuzu* 'to set a fish trap', *qalha-qa~qriuq* 'a thief' < *qriuq* 'to steal'. Rightward reduplication (Chang 1998), sometimes called suffixal reduplication (Blust 2003a, p. 194), applies to verbs of two or more syllables containing a consonant cluster in word-initial or word-medial position. It conveys repetitive or continuative aspect and might be treated as a variant of full reduplication, e.g., *angqtu~qtu* 'to think about' < *angqtu* 'to think', *mia-lundu~ndu~z* 'go in straight line (AV)' < *ma-lunduz* 'to be straight, right, correct'. CV-reduplication occurs with verbs starting with a light syllable. It conveys a repetitive meaning in motion verbs, but its semantic function in other verbs is unclear, e.g., *mu-tu~tusi* 'go there often' < *mu-tusi* 'go over there'. Triplication applies to verbs and also conveys a repetitive meaning, e.g., *ming-qa~qa~qtha* 'stop and start repeatedly' < *qtha* 'start', *makit-shka~shka~shkash* 'be gradually overwhelmed by a sense of apprehension of foreboding' < *shkash* 'to be afraid'.

Thao has no (native) compounds, but they can be formed in the creation of neologisms, for instance *matshaz=a barimbin* [fly=LNK vehicle] 'airplane' (lit. 'flying house').

#### 58.4 Word Classes

Thao distinguishes two open word classes, nouns and verbs, and the following closed classes: pronouns, noun phrase markers, demonstratives, prepositions, adverbs, negators, clausal and interclausal elements (including coordinating and subordinating conjunctions and ligatures), exclamations, and interjections. There are no auxiliaries or adjectives, the latter of which are encoded as stative verbs. Nouns are not marked for gender or number and can be divided into three main categories: (i) common nouns, (ii) locative and temporal nouns, and (iii) personal nouns (including kinship terms and given names). The major distinction between these three categories of nouns lies in the possibility or impossibility of their co-occurrence with different noun phrase markers. While common nouns can be preceded by *sa* or *na*, personal nouns are preceded by *ti*. Locative and temporal nouns, which function on the syntactic level as locative or temporal adjuncts, cannot be preceded by any noun phrase marker.

- (1) a. *pintata ti/\*sa<sup>6</sup> puni sa/\*ti bailu.*  
 cook PN/\*CN Puni CN/\*PN bean  
 ‘Puni is cooking beans.’
- b. *yaku shi-tuzi=za Ø tilha Qariawan.*  
 1SG.NEUT went-there=COS Ø yesterday Puli  
 ‘I went to Puli yesterday.’ (Blust 2003a, p. 1025)
- c. *a=lha-kayza=ihu a=m-u-tusi Ø Qariawan?*  
 IRR=LHA-when=2SG.NEUT IRR=AV-go-there Ø Puli  
 ‘When will you go to Puli?’

Verbs can be primarily divided into dynamic and stative verbs. Dynamic verbs are marked by ‘AV’  $\langle um \rangle$  or one of the following allomorphs: *m-*, *ma-*,  $\langle m \rangle$ ,  $\langle un \rangle$ ,  $\langle ung \rangle$ , and  $\emptyset$ , which are phonologically or morphologically conditioned (see Blust 1998, 2003a), e.g., *m-usha* ‘go (AV)’, *ma-tunaw* ‘win (AV)’, *k<m>ari* ‘dig (AV)’, *q<um>pit* ‘pinch (AV)’, *k<un>taq* ‘eat raw (AV)’, *sh<ung>kash* ‘fear (AV)’, *panaq* ‘to shoot’. Notice that the point of articulation of the nasal in  $\langle um \rangle \sim \langle un \rangle \sim \langle ung \rangle$  agrees with that of the following obstruent. Stative verbs are marked by *ma-* ‘STAT’ or  $\emptyset$ , e.g., *ma-qitan* ‘be good’ (see Zeitoun & Huang 2000).

## 58.5 Syntax

### 58.5.1 An Overview of Basic Clause Structure

Thao exhibits two different word orders. The first is the same as that found in other Formosan languages, i.e., the predicate—whether it is a noun, a prepositional phrase, or a verb—occurs in sentence-initial position, as in (2a–c). There are no restrictions on the occurrence of the nominal arguments, and both VOS and VSO are grammatical.

- (2) a. *Tuba yaku.*  
 Tuba 1SG.NOM  
 ‘I am Tuba.’
- b. *i-taun nak=a binanaw’az.*  
 LOC-house 1SG.GEN=LNK woman  
 ‘My wife is at home/in the house.’ (Wang 2004, p. 176, after Blust 2003a, p. 978)

6 The noun phrase marker *sa* may cliticize to the previous morpheme, as =s ‘CN’.

- c. *a=m-usha=yza yaku.*  
 IRR=AV-go=COS 1SG.NEUT  
 'I am leaving right away.'

In (3a–b), the subject occurs in sentence-initial position. In other words, Thao also exhibits an SVO word order, perhaps as a result of the influence of Taiwanese Southern Min.

- (3) a. *haya manu=a azazak?*  
 that who=LNK child  
 'Whose child is that?'
- b. *tantuqash ayuzi=a minlhafut mat yakin i-taun.*  
 elder.sibling man, male=LNK sibling CONJ 1SG.OBL LOC-house  
 'My elder brother and I are at home.'
- c. *yaku m-u-buhat itiza m-u-taun.*  
 1SG.NEUT AV-go-field return AV-go-house  
 'I worked in the fields and then returned home.' (Blust 2003a, p. 426)

The same negators, *ani* and *antu* '(do/did) not', are used in nominal and verbal clauses and may occur in sentence-initial or sentence-medial position, before the predicate, as shown in (4) and (5). There is thus no test to distinguish, a priori, nominal, prepositional, and verbal clauses.

- (4) a. *ani/antu yaku tu thaw.*  
 NEG/NEG:LNK 1SG.NEUT LNK Thao  
 'I am not Thao.' (Wang 2004, p. 262, after Blust 2003a, p. 295)
- b. *ani/antu a=i-taun nak=a apu.*  
 NEG/NEG:LNK IRR=LOC-house 1SG.GEN=LNK grandfather  
 'My grandfather will not be at home.' (Wang 2004, p. 264)
- c. *ani/antu a=makán ti ama.*  
 NEG/NEG:LNK IRR=AV:eat PN father  
 'Father will not eat.' (Youmehim Chen 2000, p. 113)
- (5) a. *yaku ani/antu thaw.*  
 1SG.NEUT NEG/NEG:LNK Thao  
 'I am not Thao.' (Wang 2004, p. 263)

- b. *yaku ani/antu a=i-taun.*  
 1SG.NEUT NEG/NEG:LNK IRR=LOC-house  
 'I will not be at home.' (Wang 2004, p. 264, after Blust 2003a, p. 446)
- c. *thithu ani/antu makaza sa but.*  
 3SG.NEUT NEG/NEG:LNK STAT:fine CN body  
 'He does not feel well.' (Blust 2003a, p. 463)

Thao displays a complex voice system, whereby the semantic role of the subject is encoded on the verb by a voice affix. Thao verbal morphology distinguishes actor voice (AV), marked by *<um>* (and its allomorphs), with the actor selected as subject (6a), patient voice (UVP), marked by *-in*, with the patient as subject (6b), and locative voice (UVL), marked by *-an*, with a patient or a location as subject (6c).

- (6) Blust (2003a)
- a. *yaku k<m>ashkash tamakuan.*  
 1SG.NEUT <AV>hoe garden  
 'I am hoeing the garden.' (p. 454)
- b. *inai=a hulus f<in>ariw-in suma.*  
 this=LNK clothes <PFV>buy-UVP other  
 'Someone has bought this shirt.' (p. 376)
- c. *kinsapiz-an thithu=a ina fukish.*  
 braid-UVL 3SG=LNK mother hair  
 'The mother braided (her daughter's) hair.' (p. 471)

### 58.5.2 Noun Phrase Structure

In Thao, a simple noun phrase consists of a pronoun, as in (3c), or a bare noun, as in (2b). A noun may be preceded by the personal marker *ti*, e.g., *ti ama* 'father' or a personal pronoun that denotes the possessor, which is followed by a linker (e.g., *nak=a apu* 'my grandfather'). Though Thao must have had case markers (Li 2011 p. 5), prenominal markers (which are optional) should rather be treated as noun phrase markers,<sup>7</sup> because there is syncretism in terms of case relations, i.e., *sa*, *na*, and *ti* can occur in nominative, accusative, and/or genitive position,

7 Wang (2004, pp. 311–328) treats *na*, *ti*, *sa*, *tu*, and *ya* all as determiners "that optionally precede noun phrases in Thao". The morpheme *tu* might better be treated as a linker, since it usually occurs between a negator and a verb, as shown in (4) above, though it might also function as

as shown by Chen (2014, pp. 20–24). An instance of such neutralization is illustrated in (7) with *ti*, which precedes the subject in (7a), the object in (7b), and the non-subject actor in (7c).

- (7) a. *ti ina m-lalas qaripuhut=a shapa.*  
 PN mother AV-peel sponge.gourd=LNK skin  
 ‘Mother is peeling sponge gourd.’
- b. *a=lhay yaku ti Shawi latata=wa patashan.*  
 IRR=give 1SG.NEUT PN Shawi one=LNK book  
 ‘I will give Shawi a book.’ (Jean 2018, p. 68)
- c. *kawi taqtaq-in ti ama.*  
 wood chop-UVF PN father  
 ‘Father is chopping wood.’ (Blust 2003a, p. 971)

It was shown in (1) that *ti* is a personal-noun marker, and *sa* and *na* are common-noun markers. This explains why *ti* and *sa*, for instance, may co-occur. In (8), the personal noun *Kilash* is preceded by a personal marker *ti*, but the whole noun phrase *Kilash=a taun* ‘Kilash’s house’ refers to the common noun *taun* ‘house’, so it is preceded by *sa*, as in (8a). Note that *ti* cannot occur before *sa*, as shown by the ungrammaticality of (8b).

- (8) Thao *e*-dictionary
- a. *i-sáy sa nak=a taun miabariz sa ti*  
 LOC-here CN 1SG.GEN=LNK house neighboring CN PN  
*Kilash=a taun.*  
 Kilash=LNK house  
 ‘My house is here, besides Kilash’s house.’

---

an emphatic marker, as in (i). The morpheme *ya* ‘when, if’ is an interclausal connector, but in (ii), its grammatical function is unclear.

- (i) *tu sa suma=wa aniamin.*  
 EMPH TOP other=LNK things  
 ‘It is someone else’s stuff.’
- (ii) *pia-kanun=uan=iza ya pania’an!*  
 CAUS.STAT-spicy=please=COS YA food  
 ‘Please make the food spicy!’

Note that *sa* also introduces a cleft sentence, as shown in (iii).

- (iii) *yaku sa [k<m><in>ay-tunu sa izáy=a azazak].*  
 1SG.NEUT CN <AV><PFV>hit-hit CN that=LNK child  
 ‘It is me who beat that child.’

- b.\* *i-sáy sa nak=a taun miabariz ti sa*  
 LOC-here CN 1SG.GEN=LNK house neighboring PN CN  
*kilash=a taun.*  
 Kilash=LNK house  
 'My house is here, besides Kilash's house.'

Noun phrases can be modified by a demonstrative (9a), a noun (9b), a numeral (9c), or a nominalized verb (9d–e). The modifier usually precedes the head noun and is followed by a linker, =a or =wa, if the modifier ends with /a/.<sup>8</sup> Note that there is an obligatory concord between the verbal modifier and the noun that it modifies; in other words, relativization, indicated in square brackets in (9d–e), is achieved through nominalization, and there is thus no specific device to encode relativization (Chen 2014). Furthermore, the semantic role of the head noun (which represents the subject of the relative) must be explicitly encoded on the modifying verb. If the head noun refers to an actor, the verb must be marked by <um> (or one of its allomorphs), as in (9d); if it denotes a patient, the verb may be marked by <in>/-in or -an (9e).

- (9) a. *haya=wa binanau'az lquzan.*  
 that=LNK woman pregnant  
 'That woman is pregnant.' (Blust 2003a, p. 202)
- b. *yaku k<m>ishkish fafuy=a kupur.*  
 1SG.NEUT <AV>scrape pig=LNK fur  
 'I cut the pig's fur.' (Blust 2003a, p. 474)
- c. *yaku ya malhkakrikriw yanan latusha=wa thau*  
 1SG.NEUT when/if AV:work have two=LNK person  
*kaidá mindahip.*  
 join.in.to.help AV:help  
 'When I (was) work(ing) there, there were two people (who came) to help me.' (Blust 2003a, p. 433)

8 There are two things to note. First, there are instances in which *haya* 'that' is not followed by the linker =wa, as in (i):

(i) Blust (2003a, p. 196)  
*haya Ø thau s<m>a-sa-sa-sas qilha lhay thithu.*  
 that Ø person <AV>RED~RED~RED-deliver wine give 3SG.NEUT  
 'That person delivers wine to him (on a regular basis).'

Second, a relative clause may follow the head noun that it modifies, though much less fre-



- d. [*t<m>ala sa kawi=a*]    *ayuzi qirqir-an qlhuran*.  
 <AV>chop CN wood=LNK man bite-UVL snake  
 ‘The man who chopped wood was bitten by a snake.’
- e. [*qirqir-an qlhuran izáy=a*]    *ayuzi i-taun pangqa*.  
 bite-UVL snake that=LNK man LOC-home rest  
 ‘The man bitten by a snake is resting at home.’

There are at least four coordinators in Thao, including *masa*, *numa* (10a), *lash* (10b), and *mat* (10c). The morpheme *mat* is a comitative marker, while both *masa* and *numa* coordinate not only noun phrases but also verb phrases and clauses (§ 58.6.4). Note that *lash* is the only coordinator that can occur with a personal noun or a kinship term.

- (10) a. *baruku masa bakung numa kakthi az’az*  
 small.bowl CONJ big.bowl CONJ chopsticks all  
*sh<in>inaw-an=iza*.  
 <PFV>wash-UVL=COS  
 ‘The bowls and chopsticks have already been washed.’ (Jean 2018, p. 199)
- b. *ama lash ti ina maka-runu sa pazay*.  
 father CONJ PN mother AV:pound-mortar CN unhusked.rice  
 ‘Father and mother are pounding rice.’ (Jean 2018, p. 198)
- c. *yaku mat ihun a=ma-parfu*.  
 1SG.NEUT CONJ 2SG.NEUT IRR=AV.RECP-fight  
 ‘I will fight with you.’ (Blust 2003, p. 594)

### 58.5.3 Personal Pronouns

The pronominal system of Thao is complex because there are case syncretism and gaps.<sup>9</sup> As illustrated in Table 58.5, there are four sets of free pronouns—neutral, oblique, dative, and genitive—and two bound pronominal forms in

quently, as in (ii); this order is not reported for any other type of modifier.

(ii) *itiza=s izáy=a binaw’az [k<m><in>ari buna]*.  
 return=CN that=LNK woman <AV><PFV>dig sweet.potato  
 ‘The woman who dug up sweet potatoes returned.’

9 The analysis presented in this section departs (in terms of the terminology adopted and the

the first-person singular (an enclitic and a suffix) marked as nominative and genitive (non-subject actor), respectively.

TABLE 58.5 Thao personal pronouns

	Free pronouns				Bound pronouns	
	NEUT SUBJ NSA	OBL	DAT	GEN	NOM	GEN-NSA
1SG	<i>yaku</i>	<i>yakin</i>	<i>nakin</i>	<i>nak</i>	=wak	-k <sup>10</sup>
2SG	<i>ihu, yuhu, uhu</i>	<i>ihun, yuhun, uhun</i>	<i>mihun</i>	<i>mihu, nuhu</i>	—	—
3SG	<i>thithu</i>	<i>thithu</i>	<i>thithun</i>	<i>thithu</i>	—	—
1PL	INCL <i>ita</i>	<i>itan</i>	—	—	—	—
	EXCL <i>yamin</i>	<i>yamin</i>	—	—	—	—
	— <i>namin</i>	—	—	—	—	—
2PL	<i>maniun</i>	<i>maniun</i>	—	—	—	—
3PL	<i>thaythuy</i>	<i>thaythuy</i>	—	—	—	—

Neutral pronouns may denote the subject, as in (11a), or the non-subject actor, as in (11b).

- (11) a. *yaku miarain k<m>an fizfiz.*  
1SG.NEUT AV:often <AV>eat banana  
'I often eat bananas.'
- b. *haya=wa apuy in-iup-an=iza yaku.*  
that=LNK fire PFV-blow-UVL=COS 1SG.NEUT  
'I have been blowing on the fire.' (Blust 2003a, p. 427)

distinctions of forms) from previous studies by Blust (2003a), Wang (2004, pp. 188–189), Li (2011, p. 7), and Jean (2018, pp. 72–77), but allows us to account for all of the forms in the first-person singular.

10 Blust (2003a, pp. 92, 96, 207) treats *-ak* and *-ik*—which represent the blending of *-an* ‘UVL’ and *-in* ‘UVP’ with the bound pronominal form *-k*—as having two allomorphs each, *-ak* and *-k*, on the one hand, *-ik* and *-k*, on the other, with accent shift. But it seems more economical to posit *-k* as a genitive (non-subject-actor) pronoun, which, by definition, must occur on UV-marked verbs.

There are three things to note: (i) we only found rare examples with *namin* '1PL.EXCL.NEUT' encoding the non-subject actor, as in (12); no example was found in which it marks the subject or possession; (ii) the dual functions of the first-person pronoun *yaku* '1SG.NEUT' correspond to the bound pronouns *=wak* '1SG.NOM', as in (13a), and *-k* '1SG.GEN', preceded by *-i(n)* 'UVP' and *-a(n)* 'UVL', respectively, as shown in (13b–c); (iii) there is partial case syncretism for the first- and second-person plural pronouns and for the third-person singular and plural pronouns. An illustration of such neutralization is given in (14a–c).

(12) Blust (2003a, p. 390)

*huya takith qaqutilh-in atu ish-funuq na takith, numa*  
that muntjac chase-UVP dog trap-mud CN muntjac then  
*sakp-in namin.*

catch-UVP 1PL.INCL.NEUT

'That barking deer was being chased by a dog, and it got bogged down in a mud patch, so we caught it.'

(13) a. *a=ma-rauz=wak.*

IRR=AV-swim=1SG.NOM

'I'll swim.' (Blust 2003a, p. 827)

b. *nak=a ama s<m>apuk rusaw kan-i-k.*

1SG.GEN=LNK father <AV>catch fish eat-UVP-1SG.GEN

'My father catches fish for me to eat (lit. 'to be eaten by me').'

c. *a=shkahul-a-k ihu u-tantu-an m-ara.*

IRR=send-UVL-1SG.GEN 2SG.NEUT go-there-UVL AV-take

'I'll send you to take it.' (Lit. 'You will be sent there to take it by me.')

(14) Blust (2003a)

a. *thaythuy a=ma-thanup atu.*

3PL.NEUT IRR=AV-bury dog

'They will bury the dog.' (p. 341)

b. *ti ama m<in>-dahip thaythuy.*

PN father <PFV>AV-help 3PL.NEUT

'Father helped them.' (p. 345)

c. *thaythuy=a hulus ma-talhin-thakthak.*

3PL.NEUT=LNK clothes AV-fall-knock.down

'Their clothes (on the line) are falling down.' (p. 339)

Oblique free forms are used when the undergoer is not the subject in a clause, as in (15a). If the undergoer is a recipient, it might be encoded by the dative (but note the gap in this pronominal set), as in (15b). Genitive (free) pronouns mark the possessor, as in (15c).

- (15) a. *sh<m>aktu ihu yakin?*  
 <AV>see 2SG.NEUT 1SG.OBL  
 'Can you see me?'  
 b. *ama a=ma-kalawa sa taun a=lha nakin/mihun.*  
 father IRR=AV-build CN house IRR=give 1SG.DAT/2SG.DAT  
 'Father will build a house for me/you.' (Blust 2003a, p. 439)  
 c. *nak=a rikus ma-kalunhan, paishish=uan uhu!*  
 1SG.GEN=LNK back STAT-itchy rub.IMP.AV=please 2SG.NEUT  
 'My back is itchy (so) please rub it (with something as a cloth soaked in hot water)! (Blust 2003a, p. 424)

#### 58.5.4 Voice, Aspect, and Mood System

The voice system of Thao was briefly introduced in §58.5.1, with illustrative examples given in (6). What is important to note is that compared to other Formosan languages, it is a reduced system in terms of voice distinctions. There is a dichotomy between actor voice (AV) and undergoer voice (UV), which subsumes UVP and UVL, but no circumstantial voice (UVC) has ever been reported. The expected marker of UVC should be *shi-* or *sha-*, and while there is in fact a prefix *shi-*, it is a verbalizer, as in *shi-buhat* 'go to work in the fields' (< *buha* 'fields') and *shi-dawaz* 'fish with a fishnet' (< *dawaz* 'fishnet'), or it encodes past events, as in (16).

- (16) Li (2011, p. 12)  
*sa izáy shinshi antu shi-da~dauk.*  
 CN that shaman no PST-RED~leisure  
 'The shaman had no free time.'

The Thao voice system is also impoverished in terms of the affix forms that are found. First, in many Formosan languages, there is a distinction between the indicative—encoded in Thao by AV <*um*>, UVP *-in*, and UVL *-an* (or *-n*, if the verb ends in a vowel)—and the non-indicative mood. The suffixes *-a*, *-aw*, *-ay*, *-ani*, and *-anay*, which are usually found in the non-indicative in other languages, are not attested in Thao. Only the non-indicative suffix *-i* is found

on UVP or UVL verbs marked as affirmative or negative imperative, e.g., *ara-i=za!* [take-IMP.UVP=COS] ‘Take it!’, *ata tu fariw-í!* [NEG.IMP LNK buy-IMP.UVP] ‘Don’t buy it!’.<sup>11</sup> Second, there is no voice marking distinction in declarative affirmative and negative clauses, as shown in (4c)–(5c).

Three other things need to be mentioned. First, there is an obvious overlap between UVP and UVL in terms of subject choice—in both cases, the subject may be the patient, as shown in (17a–b) and (17b–c)—but as mentioned by Blust (2003a, p. 239), “the basis for choosing between suffixation [with UVL or UVP] is sometimes quite obscure”, and the distinction between UVP and UVL in Thao thus requires further investigation.

- (17) a. *fafuy pa-kan-in thithu.*  
 pig CAUS-eat-UVP 3SG.NEUT  
 ‘She is feeding the pigs.’ (Blust 2003a, p. 446)
- b. *sazum qdup-in thaythuy numa uka munáy.*  
 water block-UVP 3PL.NEUT then not AV:come.here  
 ‘The water was dammed by them so that it couldn’t come over.’ (Blust 2003a, p. 790)
- c. *haya=wa atu antu nak mintahaw miku unay*  
 that=LNK dog NEG:LNK 1SG.GEN AV.expect want.to go.here  
*maqa pa-ka~kan-a-k afu.*  
 because CAUS-RED~eat-UVL-1SG.GEN rice  
 ‘That dog isn’t mine, but it got used to coming to me because I fed it rice.’ (Blust 2003a, p. 445)
- d. *numa qdup-an pitaw.*  
 then close-UVL door  
 ‘Then the door was closed.’

Second, the infix <in> is conventionally treated as a perfective marker,<sup>12</sup> which can co-occur with AV <um>, e.g., *q<m><in>riu’* [<AV><PFV>steal] ‘stole’ (Blust

11 As pointed out by Blust (2003a, p. 236), an undergoer-marked verb may also be used to mark imperative, e.g., *kawar-in thithu!* ‘Hook him!’

12 As in many other Formosan languages, the infix <in> is inserted right after the initial consonant of the verb stem or right after the infix <(u)m>, while *in-* is added to a verb stem beginning with a vowel. The prefix *in-* may also be added to a stem beginning with a consonant, as in *ing-kahiwan* (*n > ng* by assimilation) ‘in the past’, *in-lhay-an* ‘was given’, for unknown reasons.

2003a, 515) but seldom co-occurs with *-in* (< \*-en), e.g., *f<in>ariw-in* [*<PFV>*buy-UVP] ‘bought’ (Blust 2003a, 351), *k<in>ilh-’akan-in* ‘have searched for food’ (Li 2011), *k<in>an-in* [*<PFV>*eat-UVP] ‘ate’.<sup>13</sup> It is generally assumed that the UVP is zero-marked in the perfective, i.e., the single occurrence of *<in>* (or its allomorph *in-*) in UVP clauses encodes aspect rather than voice, *s<in>apuk* ‘were caught’, *lh<in>iklhik* ‘were sawed’, *p<in>asiz* ‘were wedged’, *q<in>aqutilh* ‘were chased’ (Li 2011). An alternative analysis, however, would be to treat *<in>* as a portmanteau morpheme encoding both aspect and voice. This claim is all the more plausible given that *<in>* is also a nominalizer.

Third, voice interacts closely with aspect and mood. We saw in the foregoing paragraphs that there is a mood distinction between the indicative and the non-indicative mood. The indicative mood further distinguishes between realis and irrealis (Huang 2000, p. 101, Wang 2004, p. 210). The realis mood indicates a factual event or a state in existence, while irrealis indicates an event that may take place in the future. In the realis, there is a dichotomy between the imperfective, which is unmarked, e.g., *m-ara* ‘take, is taking’ (see also (11a), (13b), (14c) above), and the perfective, marked by *<in>/in-*, e.g., *m-in-ara* [*AV-PFV*-take] ‘took’ (see also (14b)). The irrealis is marked by *a=*, e.g., *a=ma-’alah’a* [*IRR=AV*-happy] ‘will be happy’, *kih-a=kan-in* [*seek-IRR=eat-UV*] ‘will seek food’ (see also (13a), (14a), (15b) above). The full paradigm is summarized in Table 58.6 with the verb *laral* ‘take’ (Blust 2003, p. 304).

TABLE 58.6 Thao voice, aspect, and mood

			AV		UVP		UVL	
			Form	Example	Form	Example	Form	Example
IND	REAL	PFV	M <sup>14</sup> <i>&lt;in&gt;</i>	<i>m-in-ara</i>	<i>&lt;in&gt;/in-</i>	<i>in-ara</i>	<i>&lt;in&gt;...an</i>	<i>in-ara-an</i>
			M- <i>in-</i>					
		IPFV	M	<i>m-ara</i>	<i>-in</i>	<i>ara-n</i>	<i>-an</i>	<i>ara-an</i>
		IRR	<i>a=M</i>	<i>a=m-ara</i>	<i>a=...-in</i>	<i>a=ara-n</i>	<i>a=...-an</i>	<i>a=ara-an</i>
NIND	AFF	IMP	Ø	<i>ara</i>	<i>-i</i>	<i>ara-i</i>	<i>-i</i>	<i>ara-i</i>
	NEG							

13 The innovative UVP forms with *<in>...-in* in Thao are extremely rare.  
14 We follow Ross (2009) in referring to the AV infix *<um>* as M.

There are other ways to encode aspect. The clitic *=iza* (and its allomorphs *=yza ~ =za*) indicates a change of state, as in (18):

- (18) *numa m-ungqza=yza yamin i-nay.*  
 then AV-move=COS 1EXCL.NEUT LOC-here  
 ‘Then we moved here.’

When co-occurring with the irrealis marker *a=*, *=iza* indicates an event that is about to happen:

- (19) *a=m-usha=yza yaku.*  
 IRR=AV-go=COS 1SG.NEUT  
 ‘I am leaving right away.’

The clitic *=uan ~ =wan* indicates an ongoing action or state, as in (20):

- (20) Li (2011)  
 a. *miāfālhith=uan thaythuy=a minlhafut.*  
 large.family=still 3PL=LNK siblings  
 ‘They were still people of a large family.’ (p. 100, 26.3)<sup>15</sup>  
 b. *i-tusi=wan taringkuan lhkabuzun ma-ra’in=a taun.*  
 LOC-there=still place.name name STAT-big=LNK house  
 ‘(When we) were still living in Taringkuan, Lhkabuzun (had) a big house.’ (p. 100, 26.2)

We showed in §58.3.2 that the continuous/repetitive aspect is indicated by reduplication of verb root.

The imperative is zero-marked in actor-voice construction, as in (21a), and marked by the suffix *-i* in UV clauses, as in (21b). (21b) is considered more polite than (21a), though the most polite form implies the use of the verbal enclitic *=uan ~ =wan*, as shown in (22a–b).

- (21) a. *qdup pitaw!*  
 close.IMP.AV door  
 ‘Close the door!’

15 The first number refers to the text and the second to the line, thus 26.3 means Text 26, line 3.

- b. *qdup-i pitaw!*  
close-IMP.UVP door  
'Close the door!'

- (22) a. *u-shuqish=uan!*  
go-return.IMP.AV=please  
'Please return!'

- b. *ana-i=wan mihu=a qafay*  
bring-IMP.UVP=please 2SG.GEN=LNK bamboo.basket  
*a=shimul-i-k!*  
IRR=borrow-UVP-1SG.GEN  
'Please lend me your basket!' (Blust 2003a, p. 296)

### 58.5.5 Negation

Thao has the following six negators: *ani* and *antu* 'do/did (not)', as in (23a–b), *niwan* 'not yet' and *niza* 'not anymore', as in (23c–d), *uka* 'not have', as in (23e), and *ata* 'don't', as in (23f). Note that *antu* is a contraction of *ani* and the linker *tu*, which might have been borrowed from Bunun; *niwan* and *niza* are contractions of *ni* (perhaps also borrowed from Bunun) and *=wan* 'still' or *=(i)za* 'already'.

- (23) a. *ani yaku sa azazak.*  
NEG 1SG.NEUT CN child  
'I do not want children.'
- b. *Kilash antu k<m><in>an rusaw.*  
Kilash NEG:LNK <AV><PFV>eat fish  
'Kilash did not eat fish.'
- c. *niwan yaku tu k<m><in>an.*  
not.yet 1SG.NEUT LNK <AV><PFV>eat  
'I have not eaten yet.'
- d. *niza yaku tu a=munanay.*  
no.more 1SG.NEUT LNK IRR=AV:RED:come.here  
'I will not come anymore.'
- e. *uka sa nak=a azazak.*  
not.have CN 1SG.GEN=LNK child  
'I have no child.'



- f. *ata tu thanit.*  
 NEG.IMP LNK cry[AV]  
 'Don't cry!'

### 58.5.6 Interrogatives

Thao has the following interrogatives: *tima* 'who/whom' (24a–b), *numa* 'what' (24c), *kayza* / *a=lha-kayza* 'when' (24d–e), *i-na-ntua* 'where' (24f), *la-piza* 'how many (human)' (24g), *la-kuzā* 'how many (nonhuman)' (24h), *minu* 'why' (24i), and *mia-kuzā/pia-kuzā* 'how' (24j–k). They usually occur sentence-initially as the predicate, even though their functions differ: *tima* 'who/whom' and *numa* 'what' are nominal interrogatives; *kayza* / *a=lha-kayza* 'when' are adverbial interrogatives; *i-na-ntua* 'where' is a prepositional interrogative; and *la-piza* 'how many (human)', *la-kuzā* 'how many (nonhuman)', *minu* 'why', and *miakuzā/piakuzā* 'how' are verbal interrogatives.

- (24) a. *tima sa lhay ihun tuali?*  
 who CN give 2SG.OBL money  
 'Who gave you money?'
- b. *tima sa riqaz-an nuhu?*  
 who CN see-UVL 2SG.GEN  
 'Who did you see?' (Lit. 'Who was it that was seen by you?')
- c. *numa=s kafazaq-in uhu k<m>alawa?*  
 what=CN know-UDP 2SG.GEN <AV>do  
 'What can you do?'
- d. *kayza ihu tu mu-shuqish?*  
 when.REAL 2SG.NEUT LNK AV-return  
 'When did you return?'
- e. *a=lha-kayza ya mu-tanawtu?*  
 IRR=LHA-when when/if AV.go-there  
 'When will you go there?'
- f. *i-na-ntua sa mihu=a huruy?*  
 LOC-to-where CN 2SG.GEN=LNK friend  
 'Where is your friend?'

- g. *la-piza mihu=a azazak?*  
 LA-how.many 2SG.GEN=LNK child  
 'How many children do you have?'
- h. *la-kuza mihu=a hulus?*  
 LA-how.many 2SG.GEN=LNK clothes  
 'How many articles of clothing do you have?'
- i. *minu day=s azazak qa a=ma-thanit?*  
 why QUOT?=CN child QST IRR=AV-cry  
 'Why does the child cry?'
- j. *miakuza ihu tilha?*  
 AV.how 2SG.NEUT yesterday  
 'How were you yesterday?'
- k. *piakuza-n nuhu tunún qafay?*  
 how-UIP 2SG.GEN weave basket  
 'How do you weave a basket?'

There are a few things that should be noted. First, nominal interrogatives such as *tima* 'who/whom' can stay in-situ and function as the argument of the verb, as in (25). Second, when nominal interrogatives are reduplicated, they are indefinite in meaning, as in (26).

- (25) *ma-dadu ihu tima?*  
 AV-love 2SG.NEUT who  
 'Who do you love?'
- (26) *uka yaku sa numa~numa.*  
 not 1SG.NEUT CN RED~what  
 'I do not have anything.'

Third, Formosan languages usually use the term "who" instead of "what" when they ask "What is your name?". Thao is unique in using the term *kuzan* 'how' instead, as shown in (27).

- (27) *kuzan mihu=a lhanaz?*  
 how 2SG.GEN=LNK name  
 'What is your name?' (Lit. 'How do you call your name?')

58.5.7 Transcategorial Operations

Transcategorial operations include the change of a noun into a verb through verbalization and the conversion of a verb into a noun through nominalization. Note that nouns can also be further nominalized and become more abstract.

Thao displays a large array of lexical prefixes that change the lexical category, including *kilh-* ‘to search, seek’, e.g., *kilh-sazum* ‘to search for water’ (< *sazum* ‘water’), *k<un>-* ‘eat’, e.g., *k<un>tan-saqazi* ‘eat lunch (AV)’ (< *saqazi* ‘noon’), *kin-* ‘get, pick up, collect’, e.g., *kin-rusaw* ‘to catch fish’ (< *rusaw* ‘fish’), and *m-u-* ‘to go’, as in *m-u-taun* ‘go home (AV)’. There are two things to note. First, many verbalizers are made up of a lexical prefix and an AV-voice affix, and it is difficult to determine whether these affixes are monomorphemic or bimorphemic prefixes. Second, lexical prefixes may originate from verbs. For instance, the prefix *kilh-* ‘search, seek’ is related to the verb *kilhim* ‘to seek’, and the prefix *k-* ‘eat’ (which can be reconstructed at the Proto-Austronesian level) is related to the verb *kman* ‘eat’.

We have mentioned in § 58.2.3 that nominalization is closely related to voice, with the two portmanteau morphemes, <*in*>/-*in* and -*an*. We reproduce part of Chen’s (2014, p. 80) tabular categorization of Thao nominalizers with illustrative examples of argument nominalization and abstract nouns, for which we follow Comrie & Thompson’s (1985) definition. Argument nominalization indicates the semantic role of the derived verb, including agent, patient, instrument, and location. Abstract nouns refer to nouns derived from other nouns, with nearly the same distinctions as above.

TABLE 58.7 Thao nominalization

ARGUMENT NOMINALIZATION						
FORMATIVE			NOMINALIZATION	GLOSS	BASE	GLOSS
AGENT	DYN	<i>m-</i>	<i>m-ara</i>	‘who takes’	<i>ara</i>	‘to take’
		<i>&lt;m&gt;</i> ,	<i>k&lt;m&gt;ilhim</i>	‘who searches’	<i>kilhim</i>	‘to search’
		<i>&lt;um&gt;</i>	<i>k&lt;un&gt;tir</i>	‘who pinches’	<i>ktir</i>	‘to pinch’
	STAT	<i>ma-</i>	<i>ma-fazaq</i>	‘who knows’	<i>fazaq</i>	‘know’
PATIENT	PFV	<i>in-</i>	<i>in-ara</i>	‘which was taken’	<i>ara</i>	‘to take’
		<i>&lt;in&gt;</i>	<i>k&lt;in&gt;alawa</i>	‘which was done’	<i>kalawa</i>	‘to do’
	IRR	<i>-in</i>	<i>kan-in</i>	‘what will be eaten, food’	<i> kan </i>	‘to eat’
	PFV	<i>in-...-an</i>	<i>in-ishur-an</i>	‘what was pried up’	<i>ishur</i>	‘to pry up’
<i>&lt;in&gt;...-an</i>		<i>k&lt;in&gt;alisi-an</i>	‘what was dyed’	<i> kalisi </i>	‘to dye’	
	NEUT	<i>-an</i>	<i>patash-an</i>	‘what is written, book’	<i>patash</i>	‘to write’

TABLE 58.7 Thao nominalization (cont.)

ARGUMENT NOMINALIZATION					
FORMATIVE		NOMINALIZATION GLOSS		BASE	GLOSS
INSTRUMENT	Ca-	tha~thput	'broom'	thput	'to sweep'
LOCATIVE	RED-...-an	fari~fariw-an	'shop'	fariw	'to buy'
	-an	kalhus-an	'bed'	kalhus	'sleep'

ABSTRACT NOUNS					
FORMATIVE		NOMINALIZATION GLOSS		BASE	GLOSS
PATIENT	RED-in-...-an	in~in-thaw-an	'life, lifetime'	thaw	'person'
INSTRUMENT	Ca-	fa~flhuq	'towel'	flhuq	'wash'
	Ca-...-an	sha~shqurun-an	'headrest, pillow'	shqurun	'headrest'
LOCATIVE	pash-...-an	pash-afu-an	'rice container'	afu	'cooked rice'
	-an	ian-an	'residence'	ian	'sanctuary'
TEMPORAL	a=...-in	a=k-saqazi-n	'lunch time'	saqazi	'noon'

58.5.8 Valency-Adjusting Operations

In this section, a number of valency-adjusting affixes in Thao are introduced, including causative, reflexive, and reciprocal markers. There are four causative markers in Thao: *pa-* 'CAUS' (or its allomorph *p-* on vowel-initial roots), which attaches to dynamic verb roots and increases their valency by adding the causer, as shown in (28a); *pia-*, which occurs on stative verbs and has replaced the bimorphemic *pa-ka-* found in many Formosan languages, as in (28b) (Blust 2003b); and *pi-* 'CAUS.LOC' and *pu-* 'CAUS.MVT', which mainly attach to noun bases, as in (28c–d).

(28) Blust (2003a)

- a. *yaku p-apa sa ina sa azazak.*  
1SG.NEUT CAUS-carry CN mother CN child  
'I help Mother carry the child.' (p. 298)
- b. *pia-qa<rma>~rman uhu q<m>aras!*  
CAUS-⟨RED⟩~look.bad 2SG.NEUT ⟨AV⟩fence  
'Make the fence in a sloppy way (you don't need to take time to do it well)!' (p. 778)

- c. *ya ma-nasha sa saipú pi-sain yamin*  
when/if STAT-many CN turnip CAUS.LOC-here 1PL.EXCL.NEUT  
*ma-ra'in=a kaunu...*  
STAT-big=STAT container  
'If there were plenty of turnips, we would put them in a big wooden container ...' (p. 458)
- d. *numa ya tanlhiza pu-taun-in bailu.*  
then when/if evening CAUS.MVT-house-UVF bean  
'Then when evening came, the beans were put in the house.' (p. 239)

Reflexivity, which reduces valency to a single argument, is encoded through a bound root, *[anak]*, which co-occurs with different verbal affixes, e.g., *ang-anak-in* 'by oneself (UVF)', *mia-anak* 'be able to care for oneself', and *pan-anak* 'to fall down by oneself'. Note that *[anak]* might be a loan from Bunun, as suggested by Blust (2003a, p. 292).

- (29) Li (2011)
- a. *m-awra yaku ya k<m>ashi-anak thaw=a lalawa.*  
AV-not.know 1SG.NEUT COMP <AV>do-self Thao=LNK story  
'I cannot make up a Thao story by myself.' (p. 257, 37.132)
- b. *min-an~anak maniun shinshi pingqza.*  
AV.INCH-RED~self 2PL.NEUT shaman change  
'It is you, shamans, who have changed yourselves.' (p. 300, 38.169)

The reciprocal affixes *ma-*, *mapa-*, and *mapa-Ca-*, which attach to dynamic verbs, and *ma-Ca-* and *mapa-ka-*, which attach to stative verbs, reduce the valency of the clause (see Table 58.8) and mark the reciprocal or collective, as illustrated in (30a–b). The prefix *mapa-*, which co-occurs with both dynamic and stative verbs, might be a loan from Bunun.

TABLE 58.8 Reciprocal affixes in Thao

FORMA-TIVE	RECIPROCAL FORM	GLOSS	BASE	GLOSS
<i>ma-</i>	<i>ma-panaq</i>	'fight with each other'	< <i>panaq</i>	'to fight'
	<i>ma-pandu</i>	'meet with each other by chance'	< <i>pandu</i>	'to meet by chance'
	<i>ma-pasasuqa</i>	'quarrel with each other'	< <i>pasasuqa</i>	'to scold, quarrel with'
<i>mapa-</i>	<i>mapa-filhaq</i>	'spit at one another'	< <i>filhaq</i>	'to spit'

TABLE 58.8 Reciprocal affixes in Thao (*cont.*)

FORMA-TIVE	RECIPROCAL FORM	GLOSS	BASE	GLOSS
	<i>mapa-qiaqia</i>	'mourn, howl or weep together, as at a funeral'	< <i>qiaqia</i>	'to mourn'
<i>mapa-Ca~</i>	<i>mapa-qa~qalaw</i>	'seize each other'	< <i>qalaw</i>	'to seize'
	<i>mapa-tha~thawa</i>	'laugh together'	< <i>thawa</i>	'to laugh'
	<i>mapa-qa~qarman</i>	'be bad to each other'	< <i>qarman</i>	'to be bad'
<i>mapa-ka-</i>	<i>mapa-ka-bulaw</i>	'ripen, become ripe, of many things at once'	< <i>bulaw</i>	'to ripen'
<i>ma-Ca~</i>	<i>mapa-ka-daydaz</i>	'be friendly to each other'	< <i>daydaz</i>	'to love'
	<i>ma-da~diplhaq</i>	'be muddy all over, covered with mud'	< <i>ma-diplhaq</i>	'to be muddy'
	<i>ma-la~lushkin</i>	'be high-pitched, of many voices together'	< <i>ma-lushkin</i>	'to be clear, high-pitched'

- (30) Li (2011)
- a. *mapa-kay-pathay thaw masa shlilitun.*  
RECP-hit-kill      Thao and    pygmy  
'The Thao and pygmies killed each other.' (p. 123, 31.33)
- b. *mapa-'in-lawashwash sa thaw ka-taun    pin-buhat.*  
COLL-INCH-separate    CN person build-house cultivate-field  
'The people separated, building new villages and cultivating lands.'  
(p. 118, 31.5)

58.6      Complex Sentences

This section deals briefly with complex sentences, including serial verb constructions (§ 58.6.1), complement clauses (§ 58.6.2), adverbial clauses (§ 58.6.3), and coordination (§ 58.6.4).

58.6.1    *Serial Verb Constructions*

Serial verb constructions (SVCs) consist of a sequence of two or more verbs that express a single event. As such, they share the same mood and the same polarity, whether affirmative or negative; they must also share at least one argument. The second and any subsequent verb is usually marked as AV. The first verb in Thao SVCs can consist of a motion verb, e.g., *mutusi* 'go (AV)', *munay*

‘come (AV)’, or a verb encoding an evaluative concept, e.g., *mathuaw* ‘very’ (for a detailed discussion, see Jean 2018, pp. 161–170).

(31) (Blust 2003a)

- a. *m-zay ti Ali ma-thuaw mushnaw ihun*  
 AV-say PN Ali STAT-very AV:like 2SG.NEUT  
 ‘Ali said that she likes you very much.’ (p. 229)
- b. *yaku a=m-u-tusi buhat tau’aqur.*  
 1SG.NEUT IRR=AV-go-over.there field hoe[AV]  
 ‘I am going to the fields to hoe.’ (p. 302)

### 58.6.2 Complementation

Complementation refers to a clause treated as an argument of the predicate. There is thus no dependency between the two or more verbs. Thao exhibits at least two types of strategies: (i) zero strategy (i.e., paratactic complements) and (ii) occurrence of a complementizer, as in (32b). Thao has at least two complementizers: *ya*, which is homophonous with *ya* ‘when, if’, as in (32a), and *mzay*, which has been grammaticalized from the quotative verb *mzay* ‘to say’, as in (32b).

(32) Li (2011)

- a. *matangkaktun=iza, parshian ya ma-qa~quyash*  
 finish=COS forbid COMP RECP-RED~sing  
*sha~shayla=wa quyash.*  
 RED~perform=LNK song  
 ‘When it is all over, it is forbidden to keep singing and dancing for the ritual ceremony.’ (p. 355, 39.236)
- b. *maqa a=ma-fazaq sa qali mzay=a k<m>alawa*  
 so.that IRR=STAT-know CN spirit COMP=LNK <AV>build  
*ma-ra’in=a taun.*  
 STAT-big=LNK house  
 ‘So that the spirits/gods will know that (we) are building a big house.’  
 (p. 65, 15.12)

### 58.6.3 Adverbial Clauses

Thao does not have many subordinators. *Before*- and *after*-clauses are expressed by the temporal sequentiality of the verbs, and to our knowledge, *ya* ‘when, if’ is the sole productive temporal/conditional subordinator, as in (33a–

b). Subordination is demonstrated by the fact that the subject is not repeated in the main clause, as shown in (33b).

- (33) a. *numa ya m<in>u-lalu=yza, k<m>an afu.*  
 then when AV<PFV>hold-ceremony=COS <AV>eat rice  
 'Then when the ceremony was over, (we would) eat (cooked) rice.' (Li 2011, p. 42, 8.14)
- b. *yaku ya qusaz-in ani a=musha.*  
 1SG.NEUT if rain-UVF NEG IRR=AV:go  
 'If it rains, I won't go.' (Blust 2003, p. 1057)

Causal and purpose clauses are introduced by *maqa* or *a=maqa* 'because, so that', which is a conjunction that occurs in sentence-initial or sentence-medial position.

- (34) a. *maqa ya s<m>apuk sa lhkaribush, ma-thuaw undu-an.*  
 because when <AV>catch CN wild.animal STAT-very capable-UVL  
 'Because he was very capable when he was catching wild animals.' (Li 2011, p. 128, 32.9)
- b. *haya(=a) atu m-athay, thanup-i-k a=maqa antu*  
 that(=LNK) dog AV-die bury-UVF-1SG.GEN IRR=so.that NEG:LNK  
*shazk-in.*  
 smell-UVF  
 'That dog died (and) I buried it so that it would not smell.' (Blust 2003, p. 341)

#### 58.6.4 Coordination

There are three coordinators: *masa* 'and', *numa* 'then', and *numawan* (< *numa=wan* [then=still]) 'then, therefore' (Li 2014). It was shown in §58.5.2 that both *masa* and *numa* can also coordinate two noun phrases. They differ in that *masa* coordinates two verbs or verb phrases, as in (35a–b), while *numa* coordinates two clauses, as in (36a). The coordinator *numawan*, as in (36b), appears much less frequently than *numa* in the texts that were collected by Li (2011), with 59 instances of the former as opposed to 672 occurrences of the latter.

- (35) Li (2011)
- a. *ma-fazaq mun-ruza masa pit'ia.*  
 STAT-know AV.row-boat CONJ cook  
 '(I) knew how to row a boat and cook.' (p. 33, 5.8)



- b. *ma-fazaq ma-didir pazay masa q<m>a-shishi zashuq.*  
 STAT-know AV-husk unhusked.rice CONJ <AV>QA-sift husked.rice  
 ‘(I) knew how to husk rice and sift husked rice.’ (p. 32, 5.2)

(36) Li (2011)

- a. *uka=yza sa ina ama, numa ma-kasim=a*  
 not.exist=COS CN mother father then STAT-sad= LNK  
*t<un>maza~maza.*  
 <AV>hear~RED  
 ‘Having no mother or father, then (one is) sad listening (to songs).’  
 (p. 16)
- b. *numa pusha-n yamin, numawan m-ansha sa*  
 then release-UVF 1PL.INCL.NEUT therefore AV-give.gift CN  
*funfun lhay itan.*  
 seeds give 1PL.INCL.OBL  
 ‘Then we released him, so they brought seeds to give us.’ (p. 17)

### Acknowledgments

We are indebted to the reviewers for their constructive and helpful criticism. We are also grateful to Jason Lobel for his proofreading help. Unless mentioned otherwise, the data come from the authors’ field notes.

### References

- Abe, Seiya and Sumino Niida (Eds.) 安部清哉・新居田純野（共編）（2007）. *Ishi omatsu-shi “sao-go goi 4000”—kana ga kiroku shita taiheiyō no “kiki gogon”—石阿松氏『サオ語語彙4000』—仮名が記録した太平洋の“危機語言”—[*The Thao Lexicon in 4000 Words by Kilash—An Endangered Austronesian Language in Taiwan recorded in Japanese Kana—*]. Gakushūin daigaku tōyō bunka kenkyūjo kenkyū hōkoku 學習院大學東洋文化研究所調查研究報告53 [Research Institute for Oriental Cultures, Gakushuin University, Occasional Papers No. 53]. Tōkyō 東京 [Tokyo]: Gakushūin daigaku tōyō bunka kenkyūjo 學習院大學東洋文化研究所 [Research Institute for Oriental Cultures, Gakushuin University].*
- Abe, Seiya, Yoshio Nagashima and Sumino Niida (Eds.), Shigeru Tsuchida (Supervisor) 安部清哉, 長嶋善郎, 新居田純野（編）, 土田滋（監修）（2008）. *Sao-yo (taiwan shao) Goi (eigo nihongo sakuin-tuski) sao-go kenkyū shiryō* サオ語（臺灣邵語）語

- 彙 (英語·日本語索引付) サオ語研究資料II [*Thao vocabulary with English and Japanese indices: Thao language studies II*]. Gakushūin daigaku tōyō bunka kenkyūjo kenkyū hōkoku 學習院大學東洋文化研究所調查研究報告 54 [Research Institute for Oriental Cultures, Gakushuin University, Occasional Papers No. 54].
- Adelaar, Alexander (1999). Retrieving Siraya phonology: A new spelling for a dead language. In Elizabeth Zeitoun and Paul Jen-kuei Li (Eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics*, (pp. 313–354). Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert (1996). Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35(2), 272–294.
- Blust, Robert (1998a). Some problems in Thao phonology. In Shuanfan Huang (Ed.), *Selected papers from the Second International Symposium on Languages in Taiwan*, (pp. 1–20). Taipei: The Crane Publishing Co., Ltd.
- Blust, Robert (1998b). Squib: A note on the Thao patient focus perfective. *Oceanic Linguistics* 37(2), 346–353.
- Blust, Robert (1999). Pazeh phonology and morphology. *Oceanic Linguistics* 38(2), 321–365.
- Blust, Robert (2001). Squib: Thao triplication. *Oceanic Linguistics* 40(2), 324–335.
- Blust, Robert (2003a). *Thao dictionary*. Institute of Linguistics, Academia Sinica, Language and Linguistics Monograph Series A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert (2003b). Three notes on early Austronesian morphology. *Oceanic Linguistics* 42(2), 438–478.
- Chang, Laura M. (1998). Thao reduplication. *Oceanic Linguistics* 37(2), 277–297.
- Chen, Youmehim (2000). Negation in Thao and Tsou. MA thesis. Chiayi: National Chung Cheng University.
- Chen, Yu-chuan David (2014). A study of nominalization in Thao. MA Thesis. Puli, Nantou: National Chi Nan University.
- Comrie, Bernard and Sandra Thompson (1985). Lexical nominalization. In Timothy Shopen (Ed.), *Language typology and syntactic description*, vol. 3: *Grammatical categories and the lexicon*, (pp. 349–398). Cambridge and New York, NY: Cambridge University Press.
- Huang, Lillian M. 黃美金 (2000). *Shaoyu cankao yufa* 邵語參考語法 [A reference grammar of Thao]. Taiwan nandao yuyan 4 臺灣南島語言 4 [Formosan Languages Series 4]. Taipei 臺北 [Taipei]: Yuanliu chubanshe 遠流出版社 [Yuan-Liou Publishing Co.].
- Jean, Shih-lang 簡史朗 (2018). 邵語語法概論 *Shaoyu yufa gailun* [A sketch grammar of Thao] Taiwan nandao yuyan congshu 4 臺灣南島語言叢書 4 [Series on Formosan Languages 4]. Xinbei 新北 [New Taipei]: Yuanzhuminzu weiyuanhui 原住民族委員會 [Council of Indigenous Peoples]. (2nd ed.)

- Kuo, Ching-hua 郭青華 (2008). Thao texts. In Abe, Seiya, Yoshio Nagashima and Sumino Niida (Eds.), Shigeru Tsuchida (Supervisor) (安部清哉, 長嶋善郎, 新居田純野編, 土田滋監修) (2008). *Sao-go (taiwan shao) Goi (eigo nihongo sakuin-tuski) sao-go kenkyū shiryō* サオ語 (臺灣邵語) 語彙 (英語・日本語索引付) サオ語研究資料II, pp. 261–325. [*Thao vocabulary with English and Japanese indices: Thao language studies II*, pp. 261–325]. Gakushūin daigaku tōyō bunka kenkyūjo kenkyū hōkoku 學習院大學東洋文化研究所調查研究報告54 [Research Institute for Oriental Cultures, Gakushuin University, Occasional Papers No. 54].
- Lee, Amy Pei-jung (2007). A typological study on reduplication in Formosan languages. PhD dissertation. Colchester: University of Essex.
- Lee, Amy Pei-jung (2010). Reduplication and odor in four Formosan languages. *Language and Linguistics* 11(1), 99–126.
- Lee, Amy Pei-jung (2021). An overview of olfactory expressions in Formosan languages. In Łukasz Jędrzejowski and Przemysław Staniewski (Eds.), *The linguistics of olfaction*, (pp. 251–276). Amsterdam: John Benjamins Publishing Co.
- Li, Fang-kuei, Chi-lu Chen and Mei-chün Tang 李方桂, 陳奇祿, 唐美君 (1956). *Shaoyu jilue* 邵語記略 [Notes on the Thao language]. *Guoli taiwan daxue kaogu renlei xuekan* 國立臺灣大學考古人類學刊 [*Bulletin of the Department of Archaeology and Anthropology, National Taiwan University*] 7, 23–51.
- Li, Paul Jen-kuei (1976). Thao phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 47(2), 219–244.
- Li, Paul Jen-kuei (1978). The case-marking systems of the four less known Formosan languages. *Proceedings of the Second International Conference on Austronesian Linguistics, Fascicle 1*, (pp. 569–615). Pacific Linguistics C-61. Canberra: The Australian National University.
- Li, Paul Jen-kuei (1983). Notes on Thao dialects. *Bulletin of the Department of Archaeology and Anthropology, National Taiwan Normal University* 43, 48–50.
- Li, Paul Jen-kuei 李壬癸 (2001). Shaozu de diwei—jian pingjie bai lesi de shaozu diwei shuo 邵族的地位—兼評白樂思 (Blust 1996) 的邵族地位說 [On the linguistic position of Thao—Some remarks on Blust's (1996) "Some remarks on the linguistic position of Thao"]. In 詹素娟、潘英海 (編) [Su-chuan Chan and Ying-hai Pan] (Eds.), *Pingpu zuqun yu taiwan lishi wenhua lunwenji* 平埔族群與臺灣歷史文化論文集 [*Papers presented at the Symposium on the Plains Aborigines and Taiwan History*], (pp. 165–184). Taipei 臺北 [Taipei]: Zhongyang yanjiuyuan taiwanshi yanjiusuo choubeichu 中央研究院臺灣史研究所籌備處 [Institute of Taiwan History, Preparatory Office, Academia Sinica].
- Li, Paul Jen-kuei (2011). *Thao texts and songs*. Language and Linguistics Monograph Series No. 44. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei (2013). Thao loans from Bunun. *Bulletin of Chinese Linguistics* 7(2), 225–241.

- Li, Paul Jen-kuei (2014). Conjunction in Thao. In I. Wayan Arka and N.L.K. Mas Indrawati (Eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2, (pp. 401–409). Asia-Pacific Linguistics series 013 / Studies on Austronesian Languages 002. Canberra: ANU Press.
- Li, Paul Jen-kuei (2015). The preglottalised stops in three Formosan languages. In Elizabeth Zeitoun, Stacy F. Teng and Joy. J. Wu (Eds.), *New advances in Formosan linguistics*, (pp. 39–46). Asia-Pacific Linguistics series 017 / Studies on Austronesian Languages 003. Canberra: ANU Press.
- Li, Paul Jen-kuei (2016). Verbs or adverbs in Thao. *Concentric: Studies in Linguistics* 42(1), 31–44.
- Li, Paul Jen-kuei (forthcoming). *Thao—Chinese—English dictionary*. MS.
- Li, Paul Jen-kuei and Shigeru Tsuchida (2009). Yet more Proto-Austronesian infixes. In Bethwyn Evans (Ed.), *Discovering history through language: Papers in honour of Malcolm Ross*, (pp. 345–362). Pacific Linguistics 605. Canberra: The Australian National University.
- Li, Paul Jen-kuei and Rung-shun Wu 李壬癸、吳榮順 (2003). Riyuetan shaozu de feijiyixing geyao 日月潭邵族的非祭儀性歌謠 [Non-ritual folk songs of Thao at Sun-Moon Lake] *Guoli taiwan daxue kaogu renlei xuekan* 國立臺灣大學考古人類學刊 [*Bulletin of the Department of Archaeology and Anthropology, National Taiwan University*] 60, 115–162.
- Lu, Michael Shun-chieh (2003). An Optimality Theory approach to reduplication in Formosan languages. MA thesis. Taipei: National Chengchi University.
- Niida, Sumino 新居田純野 (2007). Sao-go (Taiwan) ni okeru shōten setsuji to ni kō jutsugo kaisō サオ語 (臺灣) における焦点接辞と二項述語階層 [Focus affixes and second-order predicate hierarchy in the Thao language of Taiwan]. 塩谷亨・佐々木冠・角田三枝(編). In Toru Shionoya, Kan Sasaki and Mie Tsunoda (Eds.), *Tadō-sei no tsūgengoteki kenkyū* 他動性の通言語的研究 [*Cross-linguistic studies in transitivity*], (pp. 66–78). Tōkyō 東京 [Tokyo]: Kurusio shuppan 黒潮出版 [Kurosio Publishing Co.].
- Tseng, Chia-hsing Josh (2008). The morphophonemic alternations in Thao phonology. MA thesis. Hsinchu: National Tsing Hua University.
- Tsuchida, Shigeru 土田滋 (1989). *Sao go* サオ語 [The Thao language]. *Gengo gaku dai jiten* 言語学辞典 [*Encyclopaedia of Linguistics*] 2, 15–16. Tōkyō 東京 [Tokyo]: Sansedō 三省堂 [Sansei-do Publishing Co.].
- Wang, Shan-shan (2004). An ergative view of Thao syntax. PhD dissertation. Honolulu, HI: University of Hawai'i at Mānoa.
- Weng, Cui-xia 翁翠霞 (2000). *Zuoyu he shaoyu de shi, tai, mao xitong zhi bijiao yanjiu* 鄒語和邵語的時、態、貌系統之比較與研究 [A contrastive study of tense, mood and aspect systems in Tsou and Thao]. Shuoshi lunwen 碩士論文 [MA thesis]. Jiayi 嘉義 [Chiayi]: Guoli zhongzheng daxue 國立中正大學 [National Chung Cheng University].

- Zeitoun, Elizabeth 齊莉莎 (2022). *Shaoyu cilei ji qi jiaoxue* 邵語詞類及其教學 [*Thao word classes and L2 teaching*]. Taiwan nandao yuyan congshu 111-10 臺灣南島語言叢書 111-10 [Series on Formosan Languages 111-10]. Taipei 臺北 [Taipei]: Caituan faren yuanzhuminzu yuyan yanjiu fazhan jijinhui 財團法人原住民族語言研究發展基金會 [Foundation for the Research and Development of Indigenous Languages].
- Zeitoun, Elizabeth and Lillian M. Huang (2000). Discussion on *ka-*, an overlooked marker of verbal derivation in the Formosan languages. *Oceanic Linguistics* 39(2), 391–414.

### ***Online Resources and Open Access Data***

- Blust, Robert and Stephen Trussel (Ongoing). *Austronesian comparative dictionary*. Retrieved from [www.trussel2.com/ACD](http://www.trussel2.com/ACD).
- Council of Indigenous Peoples 原住民族委員會 (2020). *Yuanzhumin renkoushu tongji ziliao* 原住民人口數統計資料 [Census of indigenous population]. Retrieved from <https://www.cip.gov.tw/en/search-result/index.html?q=population>.
- Council of Indigenous Peoples 原住民族委員會 (2021). *Yuanzhuminzu yuyan xian-shang cidian* 原住民族語言線上辭典 [Online dictionaries of indigenous languages]. Retrieved from <https://e-dictionary.ilrldf.org.tw/>.
- Council of Indigenous Peoples and Ministry of Education 原住民族委員會和教育部 (2005). *Yuanzhuminzu yuyan shuxie xitong* 原住民族語言書寫系統 [Orthographic systems for indigenous languages]. Retrieved from <https://ilrldc.tw/research/rwview/rwsystem.php>.