



**Volume**

18 (2024)

**Pages**

194 - 215

**Received**

29 Dec. 2023

**Handle**

<https://hdl.handle.net/10125/74803>

**Online**

<http://nflrc.hawaii.edu/ldc>

**Citation**

Pittayaporn, Pittayawat, Songphan Choemprayong. 2024. *A proposal for a Thai-based Moklen orthography*. *Language Documentation & Conservation* 18: 194-215.

## A proposal for a Thai-based Moklen orthography

Pittayawat Pittayaporn

*Center of Excellence in Southeast Asian Linguistics & Department of Linguistics,  
Chulalongkorn University*

Songphan Choemprayong

*Behavioral Research and Informatics in Social Science Research Unit & Department of  
Library Science, Chulalongkorn University*

Moklen [mkm] is a severely endangered Austronesian language spoken by a sea-based population along the Andaman coast in the southern part of Thailand. As an initial attempt to document this language, we propose a Thai-based orthography of Moklen that not only captures all the phonemic distinctions but also considers the social, political, and cultural contexts of Moklen communities. The proposal is based on fieldwork data collected intermittently in Bangsak, Phang Nga Province, Thailand, between 2017 and 2021. Based on Smalley's (1959) criteria of adequacy, the proposed orthography adopts a subset of consonant and vowel symbols, as well as major orthographic rules of the Thai [tha] script. Notably, it includes a tone mark that captures a tonal distinction confirmed by this study. Moreover, it is flexible enough to allow indication of sociolinguistic variations reported in previous studies. While a formal evaluation of the proposed orthography is still needed, guidelines recommended by UNESCO are also adopted to assess several sociocultural determinants that may influence the usability and acceptance of the proposed orthography. The proposed orthography is a promising tool for the documentation of the language and has potential to contribute to efforts to preserve the cultural heritage of the Moklen community.

## 1. Introduction<sup>1</sup>

Moklen [mkm] is an Austronesian language spoken by a small population scattered along the Andaman coast in southern Thailand. Besides its closest relative Moken [mwt], its precise relationship to other languages in the family is unknown. Approximately 4,000 people identified as ethnically Moklen currently live in the Khura Buri, Thai Mueang, Takua Pa, Takua Thung, and Nong Ya Plong Districts of Phang Nga Province, and in Thalang District, Phuket Province (Arunotai 2017). However, the number of actual Moklen speakers may not exceed half of that estimate as younger generations have abruptly shifted to Thai [tha] and Southern Thai [sou].

According to the criteria for assessment of language vitality proposed by the UNESCO Ad Hoc Expert Group on Endangered Languages (2003), Moklen is a severely endangered language. By the 1990s, the younger generations were already abandoning their language (Larish 1992; 1999: 103–105). According to our field observations in 2018, it is at present only used by people in the grandparental generation and up. This situation is more serious than what was described in Premsrirat (2018a), who notes, perhaps based on data from a few decades ago, that language and identity socialization still occur in the home and community. Few people under fifty years old still speak the language, none of whom are children and teenagers. While older members of the community still tend to be bilingual in Moklen and Southern Thai, the domain of use is limited only to communication within peer-groups and families. Moreover, parents generally use Thai to interact with their children at home. Literacy is extremely low among older speakers, but younger generations can read and write Thai fluently. Because almost all literate community members are either dominant or monolingual in Thai, Moklen does not have an orthography and is never used in any official media outlets. Due to a long history of discrimination and assimilation (Hogan 1972: 223; Swastham 1982: 5–6; Larish 1992), many Moklen do not see loss of cultural heritage as a threat to their identity.

Linguistically speaking, Moklen is key to understanding several issues relating to the Southeast Asian linguistic area. First, its position within Austronesian has important implications for the study of human migration and language spread. Because it appears to be only distantly related to Austronesian languages in its vicinity, Blust (1994) and Larish (1999: 453–488) speculate that they represent an early wave of Austronesian migration into the Southeast Asian peninsula. Another issue is the convergence of languages in Mainland Southeast Asia. Larish (1999: 160–161, 363–415) and Pitayaporn (2005) have discussed typological characteristics of Moken–Moklen that resemble languages of the peninsula rather than their insular relatives. The contact situation that gave rise to these mainland features is still unclear. The most interesting issue is the origin of its tonality. Larish (1997; 1999: 166–167; 2005) reports a marginal tonal contrast in Moklen, which makes it stand out among Austronesian languages. The question is thus how the language acquired its lexical tones.

Given its endangerment and its linguistic importance, the need for documentation of Moklen is urgent. Not only is it essential to the preservation of the language but also to the stewardship of the culture, worldview, and knowledge residing in the Moklen community. More importantly, the documentation of Moklen can also help promote the sociocultural identity of this population, which could also contribute to political awareness and recognition in their residential territories. Unlike its Austronesian relatives Pattani Malay [mfa] (The Royal Society of Thailand 2010; Premsrirat & Samoh 2012; Samoh 2018) and Moken (Kraisame 2016; 2018), efforts to preserve and document Moklen are only now beginning to slowly emerge. Despite initiatives by some active members of the community to write down their language, existing materials in circulation are unfortunately fragmentary, consisting only of grammatical sketches with appended word lists (Swastham 1982; Larish 1999) or very short texts (Larish 2005). The only existing lexicographical attempts we are aware of are a Moklen–Thai–English glossary compiled by members of the Thap Tawan community at Bangsak,<sup>2</sup> which consists of approximately 600 words and 120 phrases and a similar Thai–Moklen glossary based on the variety spoken in Phuket.

<sup>1</sup> This research was part of the “Moklen-Thai-English Dictionary: A Pilot Version” project funded by the Faculty of Arts, Chulalongkorn University. We would like to express our sincere gratitude to the Moklen communities for their unwavering kindness and their invaluable data, and extend a special thanks to Narumon Arunotai and her team, who provided insights and practical advice and facilitated community connections. We are also grateful for the valuable input provided by my research assistants and graduate students. We would also like to thank our research assistant Piyapath Srisomyos for his help with the preparation of the manuscript and the accompanying audio files.

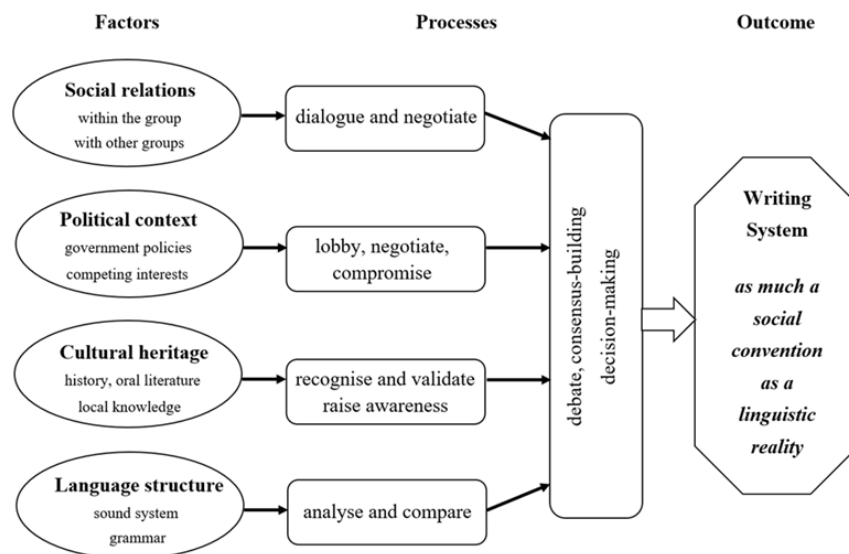
<sup>2</sup> The Moklen community at Bangsak used to spread along Bangsak beach, but they are now largely confined to the Thap Tawan, Bon Rai, and Bang Khaya communities.

One major reason for the lack of materials can be attributed to the lack of a practical orthography based on a good phonological description. The problem is clearly seen in the entries of the Thap Tawan glossary, which are rendered inconsistently and unsystematically using Thai spellings. For instance, [batɔ́j]~[matɔ́j] [029-rock.wav] ‘stone’ is spelled as บะไต, มะไต, or มาไต,<sup>3</sup> exhibiting inconsistent representation of the unstressed /a/. Similarly, [matáj] [076-dead.wav] ‘to die’ is written either as มะไต or มาตาย,<sup>4</sup> displaying an unsystematic use of long-vowel symbols and tone marks to represent the long stressed /a:/ with Tone 1. In some local schools, basic language courses are taught once a week using Thai as medium of instruction, but their success is hindered by a lack of pedagogical resources based on standard orthography designed systematically for the language. Therefore, a practical orthography is urgently needed as an important step in consistently writing this endangered language.

As part of a long-term Moken–Moklen documentation project, in this paper we report on the design of a Thai-based orthography for Moklen. After reviewing relevant literature on orthography development, we describe the sound system of Moklen based on our own fieldwork data. Subsequently, we explain the decision-making behind our proposal and discuss factors that may affect the adoption of the orthography among different Moklen communities. All orthographic forms reported here have been used in the pilot development of a Moklen–Thai–English dictionary (Pittayaporn et al. 2022).

## 2. Orthography development

Developing an orthography for an unwritten language is an art that involves balancing conflicting linguistic, sociopolitical, and practical factors (Grenoble & Whaley 2006: 137–159; Seifart 2006; Lüpke 2011; Premsrirat 2018b). As Cahill (2014) puts it, an effective orthography must be linguistically sound, acceptable to all stakeholders, and usable. To deal with such a multifaceted task, this study adopts the set of guidelines recommended by UNESCO (Robinson & Gadelii 2003), which has been applied to several orthography development initiatives (e.g., Willis Oko 2018). In addition to its comprehensiveness, it provides guiding questions that help address the multiple factors involved, as well as the processes required in decision-making. The framework is schematized in Figure 1.



**Figure 1.** Key elements in developing a writing system (reproduced from Robinson & Gadelii 2003: 13). Copyright 2003 by UNESCO. Reprinted with permission.

<sup>3</sup> In Standard Thai, these would be read as [batajɔ́], [matajɔ́], and [ma:taɔ́], respectively.

<sup>4</sup> In Standard Thai, these would be read as [matajɔ́] and [ma:taɔ́], respectively.

The first factor is social relations within the group and with other groups, which critically determine whether a proposed orthography is accepted. Because orthography is an expression of the users' identity, issues of dialect diversity, attitudes toward other languages, as well as political, religious, and ethnic ideologies need to be considered. The second factor involves political contexts. Implementing orthography is a political action that needs to consider government policies, legal compliance, the status of the language community, and all interest groups involved. The third one deals with cultural heritage. An orthography affects the way the language community relates to its cultural heritage as it provides new possibilities for the community themselves and society at large to create, access, preserve, and transmit their history, oral literature, and local knowledge. For instance, to preserve *Rong Ngeng*, a specific genre of Moklen traditional music, Ungpho (2016) found numerous challenges to the versification and restoration of the melody. It is also important to recognize, validate, and raise awareness about the effects of orthography on the roles and responsibilities of different stakeholders who create and maintain the nonwritten cultural heritage. The last factor is the linguistic dimension. Because the development of orthography would essentially regulate how the language is written and communicated symbolically, it must therefore consider language structure from various perspectives (e.g., sound system, spelling rules, and grammatical structure).

Since the above framework concerns the process of developing new writing systems, it puts more emphasis on macro-level decision-making than fine-grained linguistic features. It can be complemented by Smalley's (1959) classic and highly influential framework that focuses on the output of the design process. In Smalley's view, an adequate writing system for a minority language must strive to satisfy five functional criteria listed in Table 1 in order of importance.

**Table 1.** Criteria of adequacy proposed by Smalley (1959)

Criteria	Explanation
1 Maximum motivation	The writing system should motivate learners to write their language and must be accepted by their society and their government.
2 Maximum representation of speech	The writing system should capture all phonemic distinctions in the language.
3 Maximum ease of learning	The writing system should maximize visual contrasts between symbols and handle similar problems in consistent ways.
4 Maximum transfer	The writing system should facilitate learning to read the trade or official language.
5 Maximum ease of reproduction	The writing system should minimize effort in typing and printing.

Crucially, these criteria are sometimes in conflict, in which case the researcher must put forward a solution that balances the competing considerations. Perhaps not surprisingly, such conflicts often involve maximum representation of speech. For instance, Mal [mlf], a Mon-Khmer language spoken in Nan Province in northern Thailand, possesses an emerging contrast between a level and a rising tone (e.g., *ĩl* /po:/ 'kerosene can' vs. *ĩl̃* /pǒ:/ 'to converse'). Because of its low functional load, Filbeck (1976) notes that the tone mark <̃> for the rising tone can be omitted. This case is an example of maximum representation of speech yielding to simplicity, which addresses maximum ease of learning and maximum ease of reproduction. Another example is the case of the diphthongs in Northern Pwo Karen [pww], also known as Phlong. The plain diphthongs /ai/ and /au/ are represented by their corresponding Thai symbols <ไ-> and <เ-> in parallel with their glottalized counterparts /aiʔ/ and /auʔ/, which are written as <ไ-๕> and <เ-๕> in the orthography proposed by Cooke et al. (1976). Although this option adds two extra symbols, it avoids the introduction of final <๕> and <๕> and provides a consistent treatment of the two sets of diphthongs, achieving both maximum transfer and maximum ease of learning.

In the context of Thailand, Thai-based orthographies have become the standard since the 1960s due to the role of Thai

as the medium of instruction in school and the increasing interest in bilingual education on the part of the Thai government (Smalley 1976). In the past fifty years, Thai-based orthographies have been proposed for minority languages across the country, such as Urak Lawoi' [urk] (Hogan 1976), Chong [cog] (Premrirat 2000), Pattani Malay (The Royal Society of Thailand 2010; Premrirat & Samoh 2012), and Bisu [bzi] (Person 2001), among many others. While working separately in different places and times, the researchers follow very similar design principles. Following Smalley's tradition, Premrirat (2008; 2018b) put together a set of linguistically oriented guidelines that emphasize understanding of phonological differences between Thai and the target language (see Haas 1956 and Danvivathana 1987 for a description and analysis of the Thai writing system).

With regard to the script, they recommend that letters only be adopted as necessary. In the Thai orthography, consonant letters are grouped into three sets, namely "high," "middle," and "low," according to their effects on the interpretation of the tone marks. Phonemes that exist in both Thai and the target language should be represented by the middle and low series, unless the language is tonal. For example, the middle consonant <๓> and the low consonant <๗> should be used for /p/ and /p<sup>h</sup>/, respectively. The high consonant <๕> should be avoided.<sup>5</sup> Moreover, when the same sound can be represented by multiple letters from these two series, the most frequent and the most topographically simple ones are preferred. For example, among letters representing /t<sup>h</sup>/ in Thai, <๗> rather than <๗๑> and <๗๒> should be chosen. Importantly, for non-tonal languages, tone markers should not be used even though pitch movements can be perceived. For example, 'dog' in Nyah Kur [cbn] should be written *๗๗*, not *๗๗๑*.

On the other hand, phonemes that do not exist in Thai may be represented by modifying existing letters or assigning new values to letters that have not been used. For instance, /g/ in Pattani Malay may be written either as <๗> or <๗๑>. Similarly, if the language has a prosodic contrast, tone marks may be used to mark the distinctive categories. For example, the first tone mark <๑>, which indicates either the low or falling tone in Thai, is used to encode breathy voice in Nyah Kur and Chong.

In addition to issues concerning the script, some of the recommendations address issues related to orthographic rules. The guiding principle is for common orthographic rules in Thai to be adopted, if possible, to avoid confusion. For example, /t/ is encoded by <๓> syllable-initially, but <๓๑> at the end of the syllable. Moreover, a space may be used to indicate a word boundary like in English for ease of reading. For instance, "Think together. Do it together," may be written as *๗๗๑ ๗๗๑ ๗๗๑ ๗๗๑* rather than *๗๗๑๗๗๑๗๗๑๗๗๑*.

### 3. Phoneme inventory of Moklen

The orthography proposed is based on the Moklen dialect spoken at the Bangsak Moklen community in Bang Muang Subdistrict, Takua Pa District, Phang Nga Province. The data were collected by a team of linguistics students and researchers during six field trips in December 2017, January 2018, May 2018, October 2019, February to March 2020, and March 2021. The first three trips were part of a linguistics field methods course taught at Chulalongkorn University by the first author. The graduate students who participated in this project included Shengnan Zhang, Piyangkoon Thaweephol, Phongphorn Prayongsup, and Peerasak Sirisawat. The last two were carried out by Daniel Loss, a research assistant of the first author. In addition, supplementary data were also obtained by Pornsuda Nawarak, a semi-speaker and a teacher at a local public school. Lastly, the last trip was made by the first author and three research assistants Daniel Loss, Warunsiri Pornpotthanamas, and Lena Maluleem. The sound system of this variety generally agrees with descriptions by Swastham (1982) and Larish (1997; 1999; 2005), showing the same sets of segmental phonemes and the same patterns of variation. Crucially, our field research reveals that it has a two-way tonal contrast, confirming the hypothesis put forward by Larish (1997; 1999: 166–167; 2005).

The prosodic word in Moklen must be exactly one iambic foot. That is, it may consist of either one stressed heavy monosyllable or an unstressed light syllable followed by a stressed heavy one. Furthermore, the stressed syllable may be an open syllable with a long vowel or a closed syllable with either a short or long vowel. In contrast, the unstressed

<sup>5</sup> For Thai speakers, the middle-class consonants are psychologically the default class as they can have all the five tones. Without a tone mark, the middle class carries Tone 1, which is considered the basic tone in Thai traditional grammar. The low series when they are unmarked also has Tone 1. Both classes are more flexible in terms of tonal variation. The high class can be perceived as the marked class. Most importantly, it has Tone 5 when it is not marked for tone. Therefore, it would be more practical to write syllables in non-tonal languages with the middle class or the low class.

syllable must always be open. Interestingly, simplex words in Moklen are overwhelmingly disyllables, with very strong preference for CV.<sup>1</sup>CVC and CV.<sup>1</sup>CVVC. Table 2 provides examples of prosodic words of different shapes.

**Table 2.** Moklen prosodic words of different shapes

	(CV). <sup>1</sup> C(C)VV	(CV). <sup>1</sup> C(C)VC	(CV). <sup>1</sup> C(C)VVC
Monosyllables	/k <sup>h</sup> ú:/ [011-friend.wav] 'friend'	/nám/ [081-eat.wav] 'to eat'  /néʔ/ [137-small.wav] 'to be small'  /klák/ [003-husband.wav] 'husband'	/klá:n/ [025-bone.wav] 'bone'  /t <sup>h</sup> uát/ 'great-grandparent'  /dá:ʔ/ [153-parent's older siblings.wav] 'parent's older siblings'
Disyllables	/ʔadá:/ [006-duck.wav] 'duck'  /nac <sup>h</sup> ɛ:/ [127-book.wav] 'book'  /tuklú:/ [121-spotted dove.wav] 'spotted dove'	/dahán/ [074-very.wav] 'very'  /dábúh/ [070-boil.wav] 'to boil'  /buhún/ [117-cigarette.wav] 'cigarette'	/bəlɯəj/ [036-weary.wav] 'to be stiff'  /təŋá:k/ [084-middle.wav] 'middle'  /canát/ [013-child.wav] 'child'

Among the twenty consonantal phonemes in Moklen, only /p/, /t/, /k/, /ʔ/, /h/, /m/, /n/, /ŋ/, /j/ and /w/ can occur in syllable-final position. An interesting observation is the absence of the alveolar fricative phoneme /s/. This gap, however, is misleading as the aspirated palatal stop /c<sup>h</sup>/ is variably pronounced as [c<sup>h</sup>] or [s]. This variation applies to all onset /c<sup>h</sup>/. For instance, /c<sup>h</sup>ɛ:m/ 'Thai' can be pronounced as [c<sup>h</sup>ɛ:m] or [sɛ:m]. Similarly, /mac<sup>h</sup>ám/ 'sour' [028-sour.wav] can be realized as [mac<sup>h</sup>ám] or [masám]. This variation is also reported by Larish (1999: 143). The consonant inventory of Moklen is given in Table 3.

**Table 3.** Consonant inventory of Moklen

	Labial	Alveolar	Palatal	Velar	Glottal
<b>Voiceless stops</b>	p	t	c	k	ʔ
<b>Aspirated stops</b>	p <sup>h</sup>	t <sup>h</sup>	c <sup>h</sup>	k <sup>h</sup>	
<b>Voiced stops</b>	b	d		g	
<b>Voiceless fricatives</b>					h
<b>Nasals</b>	m	n	ɲ	ŋ	
<b>Liquid</b>		l			
<b>Glides</b>	w		j		

The most important considerations with regard to orthography relate to the four variations in syllable structure in the initial unstressed syllable of disyllables. These arise through phonological variation and are also discussed by Larish



(1999: 170–180). First, the voiced stops are realized either as stops [b] and [d] or as sonorants [m] and [l] as the onset of the initial syllable. For example, /batɕj/ [029-rock.wav] ‘stone’ can be realized as [batɕj] or [matɕj]. Similarly, /dapán/ [030-stove.wav] ‘stove’ can be pronounced as [dapán] or [lapán]. Second, the contrast between /l/ and /n/ are neutralized in the onset position of the initial syllable so that [l] and [n] are interchangeable. For example, /lekót/ [086-back.wav] ‘back’ can be pronounced as [lekót] or [nekót]. The third variation is a dialectal one. In this dialect, the voiced velar stop cannot occur as onset of the initial syllable so that words beginning with /g/ in the speech of older speakers of other dialects always have /d/ in this variety. For example, /gajáh/ ‘elephant’ in other varieties corresponds to /dajáh/ [031-elephant.wav] in Bangsak. The interaction among these three variation patterns is clearly illustrated by the variable pronunciation of /dutɕj/ ‘louse’ as [dutɕj], [lutɕj], or [nutɕj], corresponding to /gutɕj/ in more conservative varieties. Lastly, /tə/ is variably pronounced as [tə] or [kə] if the following syllable begins with another stop (cf. Larish 1999: 178). For example, /təbó:ŋ/ [033-cheek.wav] ‘cheek’ can be pronounced either as [təbó:ŋ] or [kəbó:ŋ].

Unusual for Austronesian languages, Moklen has a total of eighteen vowels in stressed syllables, comprising nine pairs of short and long ones. However, the vowel length contrast is crucially completely neutralized in the initial syllable of disyllabic words. In other words, only /a/, /i/, /u/, /ə/, and /ɛ/ can occur in unstressed syllables. The duration of these unstressed vowels is an intermediate length, between that of the short and long vowels in stressed syllables. In addition, the language also has an extra-short neutral vowel /ə/, which only occurs in unstressed syllables. Its phonetic realization also varies greatly depending on its phonetic environment. For example, the /ə/ sounds rather like [a] in /məʔé:n/ [034-bathe.wav] ‘to bathe’ because of the low vowel in the final syllable but closer to [i] and [u] in /cʰəták/ [016-bean.wav] ‘bean, nut’ and /bəcʰúj/ ‘iron’, under the influence of the palatal and labial onsets, respectively. The Moklen vowel inventory is given in Table 4.

**Table 4.** Vowel inventory of Moklen

	Front	Back	
		Unrounded	Rounded
High	i, i:	ɯ, ɯ:	u, u:
Mid-high	e, e:	ɤ, ɤ:	o, o:
Mid-low	ɛ, ɛ:	ə	ɔ, ɔ:
Low		a, a:	

Diphthongs: iə, uə, uə

Another interesting phenomenon is phonetic vowel nasalization, also reported in Larish (1999: 167–169). Any vowel that occurs before or after a nasal consonant is phonetically nasalized. For example, /ɲð:k/ [083-pound.wav] ‘to pound’ and /təŋá:k/ [084-middle.wav] ‘middle’ are phonetically realized as [ɲð̃:k] and [təŋá̃:k], respectively. In disyllables, if the onset of the initial syllable is a nasal vowel, and the onset of the final syllable is a glottal consonant, nasalization extends throughout the word. For instance, /məʔi:ʔ/ ‘sweet’ and /nəhó:j/ [075-spoon.wav] ‘spoon’ are pronounced as [məʔĩ:ʔ] and [nəhó̃:j], respectively. Despite the pervasiveness of vowel nasalization, nasality is not contrastive since it is not used to make lexical distinctions. Another pattern is the phonological variation of unstressed /ʔə-/, which is more often realized as [ʔa-] in normal speech. For example, the word for ‘head’ can be pronounced as either [ʔaták] or [ʔəták].

The most intriguing feature of Moklen is its lexical tone. Larish (1997; 1999: 166–167; 2005) states that the pitch pattern of a Moklen word is largely predictable from its syllable structure but presents a few minimal pairs that seem to show that pitch is lexically contrastive. For example, [walɕj̃] [038-day.wav] ‘day’ and [kɔla:t̃] [039-hot.wav] ‘hot’ are distinguished from [walɕj̃] [041-eight.wav] ‘eight’ and [kɔla:t̃] [040-mushroom.wav] ‘mushroom’. He therefore tentatively concludes that Moklen is tonal. Intriguingly, this hypothesis is confirmed by our fieldwork data, which reveal that Moklen has two phonemic tones on stressed syllables. In citation forms, Tone 1 is realized with a high pitch in the first half of the rime, usually with a falling contour [N] in nonchecked syllables. In contrast, Tone 2 is characterized by a mid-level pitch in the first half of the rime, often accompanied by a rise toward the end [L]. Whether the tones are

realized with the final rise or fall seems to depend on syllable structure and intonational context. Table 5 provides examples of tonal minimal pairs.

**Table 5.** Minimal pairs illustrating lexical tones in Moklen

Tone 1	Tone 2
/wal̥ɰj/ [090-left.wav] ‘left’	/wal̥ɰj/ [041-eight.wav] ‘eight’
/kɔl̥á:t/ [039-hot.wav] ‘hot’	/kɔl̥à:t/ [040-mushroom.wav] ‘mushroom’
/dal̥á:ŋ/ [172-mosquito.wav] ‘mosquito’	/dal̥à:ŋ/ [173-mud lobster.wav] ‘mud lobster’
/b̥ɰj/ [171-get.wav] ‘to get’	/b̥ɰj/ [169-2SG.wav] ‘2s pron.’
/ʔá:k/ [174-place (v).wav] ‘to place’	/ʔà:k/ [170-crow.wav] ‘crow’

Moklen can thus be considered a language with word tones, as each prosodic word can only bear one tone on the stressed syllables. The initial syllables of disyllables are always phonologically toneless. Quantitatively speaking, the tonal contrast is marginal as few minimal pairs exist. This has to do with the fact that Tone 2 is found in a much smaller number of words than Tone 1. The biggest mystery is how the tonal contrast emerged in Moklen in the first place.

#### 4. Applicability of a Thai-based writing system

Because Moklen is already severely endangered, a practical orthography of Moklen is urgently needed to document traditional knowledge before it vanishes upon language death. Rather than waiting for a full-scale orthography development project with community engagement, the current proposal is the outcome of the second author’s observations of the community’s sociopolitical dynamics and conversations with community leaders in January 2018. Considering the Moklen community’s social relations, political contacts, cultural heritage, and language structure (Robinson & Gadelii 2003), a Thai-based writing system is an obvious choice, as we show below.

With respect to social relations, Thai is the preferred language in the community at large, except for senior members. Based on the results from the interviews during our field trips, while community leaders and senior members view the Moklen language as one of the main elements for sustainability of the community, Thai is associated with upward social movement, modernity, and economic opportunity. This is particularly true for the younger generations who were reluctant to speak Moklen, particularly in public.

This complete dominance of Thai is due to two major trends in the community. The first one is the rapid increase in literacy through formal education. Although the schools near the community have attempted to provide basic Moklen training once a week, they still lack qualified instructors who are fluent speakers of the language. The second trend is the change of the demographic landscape of the community due to industrialization and urbanization. The arrival of tourism-based businesses and convenience stores as well as government agencies continue to bring people from outside the community into the area. At the same time, a number of individuals in the working-age population have left the community to seek better opportunities elsewhere, mostly in the cities. Inter-marriage also causes demographic change in the community. Thus, the increase in contact with non-Moklen speakers has accelerated the rise of Thai as the main language of everyday communication within the community.

As for political contexts, the revitalization of minority languages has been increasingly recognized as significant by the



Thai government. As one of the member states of the United Nations, Thailand must be in compliance with the UN's (2007: Article 13) Declaration on the Rights of the Indigenous People, which protects the rights of Indigenous people to "revitalize, use, develop and transmit to future generations their histories, languages, oral traditions, philosophies, writing systems and literatures, and to designate and retain their own names for communities, places and persons." In 2018, the Royal Society, the Thai national authoritative agency regulating cultural and linguistic affairs, drafted a strategic plan for a national language policy (Office of the Royal Society 2018), one of whose objectives is to preserve Indigenous languages in order to safeguard the diversity of cultural identities. In response to the need to develop a writing system for unwritten Indigenous languages, the Royal Society has also published a number of Thai-based writing-system manuals for Indigenous languages in Thailand, such as Pattani Malay (the Royal Society of Thailand 2010) and Gong [ugo] (the Royal Society of Thailand 2018).

Regarding cultural heritage, language is pivotal to the sustainability of local culture and knowledge. To our knowledge, there is currently no systematic collection or documentation of local knowledge, history, and folklore in Moklen. Most documentation available to date is mostly recorded in Thai. In some communities, such as in the villages of Thap Tawan, Thung Wa, and Lampi, where there is a strong presence of the Moklen population, Moklen cultural centers were set up to store and exhibit the stories of this ethnic group. Most centers exhibit objects used in daily life, as well as display signs in Thai narrating the local history. If needed, Moklen words are transcribed ad hoc into the Thai orthography. In addition, Phang Nga's Provincial Cultural Office has launched projects to promote and preserve the lifestyles, cultures, local wisdom, and languages of the sea people in the province. Language training for children in the Moklen communities was also included in these initiatives. In summary, these efforts demonstrate the support from Thai administrative authorities for developing a Thai-based orthography to facilitate the documentation of indigenous languages. Given the role of the Thai language and its orthography in current documentation practices, a Thai-based orthography for Moklen would allow literate community members to start writing their language and documenting their cultural heritage. Moreover, using Thai as a reference orthography would require relatively little effort on the part of the learners, especially since it is already a dominant language for the new generations of ethnic Moklen. This approach would enhance the long-term preservation of the local history and stories.

Finally, the Thai orthography can be modified to suit the linguistic structure of Moklen with relatively little difficulty. With respect to the consonants, only two phonemes, namely /ɲ/ and /g/, do not have correspondents in Thai. Fortunately, there are many symbols in the Thai script that can be readily employed to represent these two sounds. With respect to the vowels, the inventories in the two languages are identical, with the same quality and quantity contrast. As for tones, the Moklen system is much more rudimentary than Thai and can thus be easily accommodated. Moreover, the Moklen syllable structure exhibits phonotactic restrictions very similar to Thai, including obligatory onsets, sonority sequencing of onset clusters, laryngeal neutralization in the coda, and the bimoraic requirement of stressed syllables. The only potential problem is the encoding of final /ʔ/ and /h/, but this can be readily accounted for by extending existing conventions in the Thai orthography.

Unfortunately, developing an orthography based on that of the dominant language presents two important drawbacks (Seifart 2006). Firstly, an orthography adapted from that of the dominant language typically comes with inconsistencies that exist in the source writing system. An example of imperfections associated with the Thai orthography is the incompatibility of the vowel shortener and the tone marks. Consequently, non-high vowels /e/, /ɤ/, /ɛ/, and /ɔ/ cannot be graphically marked as short in syllables that carry one of the tone marks. For instance, it is not possible to predict if *in* is pronounced as /l̥ɛ̃j/ and /l̥ɛ̃:j/. Secondly, it is potentially less emblematic than newly created ones. On the contrary, it promotes national unity in the face of linguistic diversity. Despite these two disadvantages, we argue that the advantages outlined above demonstrate that a Thai-based orthography is still the best available choice for Moklen.

## 5. Proposed writing system

Given Moklen's structural similarities to Moken, the current proposal is adapted from Kraissame's (2016: 48–62) Moken orthography. Essentially, it is an Indic-derived abugida or alphasyllabary, which comprises a set of consonant graphemes with an inherent vowel and a set of vowel diacritics (Coulmas 2003: 145–148). For example, the letter <n> is pronounced as /kɔ:/ when it occurs alone, or as /ko/ when closed by a consonant (e.g., /koŋ/ for *nɔ*). To encode syllables with other vowels, vowel diacritics must be placed before, after, above, under, or even around the consonant symbols they modify (e.g., /ko:/ for *īn*, /ka:/ for *nɪ*, /ki:/ for *ñ*, /ku:/ for *ŋ*, and /kɤ:/ for *in*). In addition, the proposed Moklen or-

thography also makes use of a vowel shortener <◌̣> and a tone mark <◌̂>.

Following Preamsirat's recommendation (2008; 2018b), only letters from the middle and low series are selected. Even though Moklen is a language with lexical tones, its system is very rudimentary. The middle and low series alone are adequate for representing the two tones. Therefore, discarding the high consonants keeps the orthography simple for maximum ease of learning. Moreover, it is worth noting that two relatively rare letters are used to represent consonant phonemes that do not exist in modern Thai. Departing from Preamsirat (2008; 2018b) and Kraisame (2016: 57–58), <ฃ> and <ค>, which are pronounced /j/ and /k<sup>h</sup>/ in Standard Thai, are used to represent /ɲ/ and /g/ in the proposed Moklen orthography. Using these marked letters avoids diacritics that do not exist in regular Thai writing, such as <◌̣>, <◌̂>, and <◌̣̂>. Not only is such a combination visually difficult to process but also awkward when the text is underlined. Moreover, such modified consonants would be impossible to combine with superscripted vowels such as <◌̣̂> or <◌̣̂̂> in most word processors. While the choice may deviate from maximum transfer, it simultaneously satisfies maximum representation of speech, maximum ease of learning, and maximum ease of reproduction.

Another important feature of the proposed writing system is the use of the letters that are pronounced as voiced stops /b/ and /d/ syllable-initially to represent voiceless sounds in syllable-final position. Retaining the Thai convention of writing final /-p/ and /-t/ with the letters <บ> and <ต> is in line with the criteria of maximum transfer. Table 6 lists all consonant phonemes and their corresponding symbols in the proposed writing system for Moklen.

**Table 6.** Moklen consonant phonemes and their corresponding symbols

Phonemes	Symbols	Examples	Usage notes
/p/	<ป>	ปาด /pá:t/ [047-four.wav] 'four' กาแฟ /kapú:n/ [048-lime(CaO).wav] 'lime'	syllable-initially
	<บ>	กอล็อบ /kɔ̀lɔ̀:p/ [096-centipede.wav] 'centipede' แอนบาน /ʔebá:p/ [097-ancestor.wav] 'ancestor'	syllable-finally
/t/	<ต>	ตึกตัก /tekát/ [049-tick.wav] 'tick' กอดาน /kɔ̀tá:n/ [050-forest.wav] 'forest'	syllable-initially
	<ต>	ซึ้อต /c <sup>h</sup> əʔt/ [098-sweat.wav] 'sweat' เบต /bé:t/ [099-go out.wav] 'to exit'	syllable-finally
/c/	<จ>	จปาน /capá:n/ [051-wood plank.wav] 'board' กึ่จัน /kəćé:n/ [052-ring.wav] 'ring'	
/k/	<ก>	กึ่ตาน /kətá:m/ [053-crab.wav] 'crab' กลาง /klá:ŋ/ [054-hawk.wav] 'hawk' ลึ่มัก /ləmák/ [055-fat.wav] 'fat'	
/ʔ/	<อ>	อาปุ /ʔapú:j/ [056-fire.wav] 'fire' อุเอน /ʔuʔé:n/ [057-water.wav] 'water'	syllable-initially
	<ห̂>	อาทห̂ /ʔaká:ʔ/ [101-elder sibling.wav] 'older sibling' ลนทห̂ /laké:ʔ/ [102-stomach.wav] 'gut, intestine'	after a long vowel
	-	บูโษะ /buc <sup>h</sup> óʔ/ [103-rotten.wav] 'putrid' กาปะะ /kapáʔ/ [105-pit viper.wav] 'Malayan pit viper'	after a short vowel

Phonemes	Symbols	Examples	Usage notes
/p <sup>h</sup> /	<พ>	บุพ <sup>ุ</sup> /nup <sup>h</sup> lùj/ [058-buy.wav] ‘to sell’ พล <sup>ุ</sup> อ <sup>ุ</sup> /p <sup>h</sup> lò:ʔ/ [059-pleasing.wav] ‘beautiful’ น <sup>ุ</sup> พ <sup>ุ</sup> ล <sup>ุ</sup> าด /nəp <sup>h</sup> là:t/ [060-fall.wav] ‘to fall’	
/t <sup>h</sup> /	<ท>	น <sup>ุ</sup> ท <sup>ุ</sup> /nat <sup>h</sup> ám/ [061-trample.wav] ‘to trample’ ท <sup>ุ</sup> /t <sup>h</sup> áw/ [062-grandparent.wav] ‘grandparent’ ท <sup>ุ</sup> น <sup>ุ</sup> /t <sup>h</sup> əléh/ [063-swift.wav] ‘swift’	
/c <sup>h</sup> /	<ช>	ช <sup>ุ</sup> ต <sup>ุ</sup> ก /c <sup>h</sup> ták/ [016-bean.wav] ‘bean’ ม <sup>ุ</sup> ช <sup>ุ</sup> /mac <sup>h</sup> ám/ [028-sour.wav] ‘sour’ ช <sup>ุ</sup> บ <sup>ุ</sup> ง /c <sup>h</sup> ɔbáj/ [064-earring.wav] ‘earring’	
/k <sup>h</sup> /	<ก>	ท <sup>ุ</sup> ย /k <sup>h</sup> áj/ [065-smoke.wav] ‘smoke’ ม <sup>ุ</sup> ก <sup>ุ</sup> น /mak <sup>h</sup> ún/ [066-yam.wav] ‘yam’ ค <sup>ุ</sup> น <sup>ุ</sup> ด /k <sup>h</sup> aná/ [067-very, seriously.wav] ‘very’	
/b/	<บ>	บ <sup>ุ</sup> ต <sup>ุ</sup> ย /batáj/ [029-rock.wav] ‘stone’ บ <sup>ุ</sup> ก /bú:k/ [068-beneath.wav] ‘under’ ก <sup>ุ</sup> บ <sup>ุ</sup> อ <sup>ุ</sup> ก /kabók/ [069-lazy.wav] ‘lazy’	
/d/	<ด>	ด <sup>ุ</sup> บ <sup>ุ</sup> ย /dəbúh/ [070-boil.wav] ‘to boil’ ด <sup>ุ</sup> ล <sup>ุ</sup> /dalám/ [071-inside.wav] ‘in, deep’ ด <sup>ุ</sup> ค <sup>ุ</sup> /ʔadá:/ [006-duck.wav] ‘duck’	
/g/	<ง>	บ <sup>ุ</sup> ก <sup>ุ</sup> อ <sup>ุ</sup> ท <sup>ุ</sup> /bagò:ʔ/ [072-frog.wav] ‘frog, toad’ ด <sup>ุ</sup> ก <sup>ุ</sup> ม <sup>ุ</sup> /dagà:/ ‘basket’ ด <sup>ุ</sup> จ <sup>ุ</sup> น /təgú:n/ [073-thunder.wav] ‘thunder’	
/h/	<ฮ>	ด <sup>ุ</sup> ฮ <sup>ุ</sup> น /dahán/ [074-very.wav] ‘much, many’	syllable-initially
	<ฮ>	น <sup>ุ</sup> ฮ <sup>ุ</sup> ย /nəhó:j/ [075-spoon.wav] ‘spoon’ น <sup>ุ</sup> บ <sup>ุ</sup> ค <sup>ุ</sup> /bəkáh/ [106-break.wav] ‘to break’ น <sup>ุ</sup> บ <sup>ุ</sup> ย /nipíh/ [108-thin.wav] ‘thin’	syllable-finally
/m/	<ม>	ม <sup>ุ</sup> ต <sup>ุ</sup> /matáj/ [076-dead.wav] ‘to die’ น <sup>ุ</sup> ม <sup>ุ</sup> น <sup>ุ</sup> /namá:ʔ/ [077-enter.wav] ‘to enter’ ด <sup>ุ</sup> ล <sup>ุ</sup> น /dalúm/ [078-needle.wav] ‘needle’	
/n/	<น>	ม <sup>ุ</sup> น <sup>ุ</sup> อ <sup>ุ</sup> น /maʔé:n/ [079-play.wav] ‘to play’ จ <sup>ุ</sup> น <sup>ุ</sup> ด /caná:t/ [013-child.wav] ‘child’ น <sup>ุ</sup> /nám/ [080-six.wav] ‘six’	

Phonemes	Symbols	Examples	Usage notes
/ɲ/	<ญ>	จู้ว /ɲám/ [081-eat.wav] ‘to eat’ มีญาค /məɲá:t/ [082-oil.wav] ‘oil’ ญอญ /ɲò:k/ [083-pound.wav] ‘to pound’	
/ŋ/	<ง>	ตึ่งาก /təŋá:k/ [084-middle.wav] ‘middle’ มีจอง /bicó:ŋ/ [085-mouse deer.wav] ‘barking deer’	
/l/	<ล>	แลท้อค /ləkót/ [086-back.wav] ‘back’ ปาล็ย /palǎj/ [087-stringray.wav] ‘rayfish’ บุลาน /bulá:n/ [088-moon.wav] ‘moon’	
/w/	<ว>	วาล็ย /walǎj/ [090-left.wav] ‘left’ กัว /kuwáj/ [091-rope.wav] ‘cord, rattan’ วัง /wájŋ/ [092-find.wav] ‘to seek’	
/j/	<ย>	บายย /bajá:j/ [093-midwife.wav] ‘midwife’ เอ็ย /ǎj/ [019-dog.wav] ‘dog’ ดูญัก /dujù:k/ [095-seven.wav] ‘seven’	

The proposed set of consonant symbols are sufficient to capture the sociolinguistic and dialectological variation mentioned in §3. The principle is to select the symbols that represent the actual pronunciation of each word. For example, the word for ‘louse’ is pronounced [dutǎj], [lutǎj], or [nutǎj] in Bangsak but can also be pronounced [gutǎj] in some other dialects. Instead of arbitrarily choosing one spelling as standard, *ดูเต็ย*, *ลูเต็ย*, *นูเต็ย*, and *ยุเต็ย* are all acceptable in this proposed orthography.

As for vowels, the current proposal follows Thai conventions closely to avoid confusing Thai-literate members of the Moklen community. Although we attempt to achieve a one-to-one correspondence between phonemes and graphemes, we still maintain all the allography in the Thai orthography for consideration of maximum transfer. For example, the short /ɔ/ is written with <๓> in most environments, except in syllables ending with /-ʔ/ or /-h/, in which it is encoded as <๓๕>. In addition, under-differentiation is allowed in the case of /ɽ/ and /ɽ:/ because the shortener <๓> is not allowed on <๓> (Haas 1956: 71). Table 7 lists all vowel symbols in the proposed Moklen orthography.

Table 7. Moklen vowel phonemes and their corresponding symbols

Phonemes	Symbols	Examples	Usage notes
/i/	<๓>	อวจิน /ʔaɲín/ ‘wind’ กูญิต /kujít/ ‘turmeric’ นีนี้อ /nipih/ [108-thin.wav] ‘thin’	Always <๓> in unstressed syllables (vowel length neutralized)

Phonemes	Symbols	Examples	Usage notes
/i:/	<๓>	คาวจี้ /dací:/ [109-worm.wav] ‘worm’ กี้ซึน /kəc <sup>h</sup> í:n/ [110-comb.wav] ‘comb’ นื้อน /niíú:n/ [111-coconut.wav] ‘coconut’	
/u/	<๓>	นื้อน /nimúk/ [112-squid.wav] ‘squid’ นึ่ง /núŋ/ [113-steam (v).wav] ‘to steam’	
/u:/	<๓>	คั้งค /təŋú:t/ [114-smell.wav] ‘to smell’ กี้ซึน /kəhú:n/ [116-turkey berry.wav] ‘pea eggplant’	
	<๓-อ>	บั้งก้อ /páŋkú:/ ‘millipede’	In open syllables
/u/	<๓>	บุฮุน /buhún/ [117-cigarette.wav] ‘cigarette’ คี้จุน /ticúm/ [008-bird.wav] ‘bird’ คูนง /dubúŋ/ [118-roof.wav] ‘roof’	Always <๓> in unstressed syllables (vowel length neutralized)
/u:/	<๓>	กาตุง /katú:ŋ/ [119-firewood.wav] ‘firewood’ ม็ญน /məŋú:n/ [120-hide.wav] ‘to hide’ ตุกตุ /tuklú:/ [121-spotted dove.wav] ‘dove’	
/e/	<๓>	บี่ล็ด /bəlét/ [122-seed.wav] ‘pill’ ชาเป็น /c <sup>h</sup> apén/ [005-crest.wav] ‘chicken’s comb’	
	<๓-ะ>	*เนะ /néʔ/ (unattested) *เนะฮ /néh/ (unattested)	In syllables ending with /-ʔ/ or /-h/
/e:/	<๓>	คี้เทศ /tit <sup>h</sup> é:t/ [023-pumpkin.wav] ‘pumpkin’ คี้เวด /dəwé:t/ [123-money.wav] ‘money’	
/ɤ/	<๓>	ซึนทลิว /c <sup>h</sup> əp <sup>h</sup> lɤw/ [124-trousers.wav] ‘pants’	
	<๓>	เอ็ย /ʔɤj/ [019-dog.wav] ‘dog’ กี้ย็ย /kəjɤj/ [125-shark.wav] ‘shark’	In syllables ending with /-j/
	<๓-อ>	*เนอะ /nɤʔ/ (unattested) *เนอะฮ /nɤh/ (unattested)	In syllables ending with /-ʔ/ or /-h/

Phonemes	Symbols	Examples	Usage notes
/ɣ:/	<เ-อ>	ปืเทอ /pətʰɣ:/ [126-ginger.wav] ‘shampoo ginger’	
		นาเชอ /nacʰɣ:/ [127-book.wav] ‘book’	
	<เ-็>	ตีเท็ก /təkʰɣ:k/ [128-beard.wav] ‘beard’ กาเด็ง /kadɛ́:ŋ/ [129-heel.wav] ‘heel’	
	<เ->	เนย /nɣ:j/ [130-sterilised milk.wav] ‘sterilized milk’	In syllables ending with /-j/
/o/	-	ตีคัง /təkʰoŋ/ [131-maize.wav] ‘maize’	
	<เ-ะ>	โต๊ะ /tóʔ/ [132-table.wav] ‘table’ *โต๊ะฮ /tóh/ (unattested)	In syllables ending with /-ʔ/ or /-h/
/o:/	<เ->	ดาโงง /dajó:ŋ/ [133-high.wav] ‘tall’ โง้ง /ŋó:ʔ/ [134-sleepy.wav] ‘sleepy’	
/ɛ/	<เ-็>	มาเท็ด /makét/ [135-hurt.wav] ‘hurt’ กานนี่ก /kané:k/ [100-pot.wav] ‘pot’ นาเอ็ก /naʔék/ [136-go up.wav] ‘to go up’	Always <เ-> in unstressed syllables (vowel length neutralized)
	<เ-ะ>	เนะ /néʔ/ [137-small.wav] ‘small’ มาเอะฮ /maʔéh/ [138-feaces.wav] ‘to defecate’ ทีและฮ /thələ́h/ [063-swift.wav] ‘to be swift’	In syllables ending with /-ʔ/ or /-h/
/ɛ:/	<เ->	แปด /pé:t/ [045-knife.wav] ‘knife’ ออเนน /ʔoné:n/ [139-hornet.wav] ‘hornet’ เนนเตด /neté:t/ [140-tremble.wav] ‘to shiver’	
/ɔ/	<อ-็>	มาน็อก /manók/ [141-chicken.wav] ‘chicken’ นอม็อก /nomók/ [142-pierce.wav] ‘to pierce’	Always <อ> in unstressed syllables (vowel length neutralized)
	<เ-าะ>	กอละฮ /kolóh/ ‘skinny’	In syllables ending with /-ʔ/ or /-h/
/ɔ:/	<อ>	ลอกอด /ləkót/ [143-bite.wav] ‘to bite’ ดอมอก /domók/ [144-bubble.wav] ‘bubble, foam’ ลืบอง /ləbó:ŋ/ [145-bamboo shoot.wav] ‘bamboo shoot’	



Phonemes	Symbols	Examples	Usage notes
/a/	<๓>	แตกััน /tɛkán/ [146-mat.wav] ‘mat’ จู้วัด /cəwát/ [147-clothe.wav] ‘cloth’ บาจัน /bacán/ [148-shrimp paste.wav] ‘shrimp paste’	
	<๓>	กาปะ /kapáʔ/ [105-pit viper.wav] ‘Malayan pit viper’ คายะฮ์ /dayáh/ [031-elephant.wav] ‘elephant’	In syllable ending with /-ʔ/ or /-h/
	<๓>	แกคำ /kɛdám/ [149-ant.wav] ‘ant’ มาน่า /manám/ [150-plant (v).wav] ‘to plant’	Representing /am/
	<๓>	กััน /kənáj/ [017-soil.wav] ‘soil, sand’ กาถ /kakáj/ [043-foot.wav] ‘foot’	Representing /aj/
	<๓>	บี่เตา /bəláw/ [004-wife.wav] ‘wife’ ตีชา /tic <sup>h</sup> áw/ [151-tomorrow.wav] ‘tomorrow’	Representing /aw/
/a:/	<๓>	กอดาง /kɔdáj/ [152-shrimp.wav] ‘shrimp’ จานาค /canát/ [031-child.wav] ‘child’ คาห /dá:ʔ/ [153-parent’s older siblings.wav] ‘parent’s older sibling’	
/iə/	<๓>	เอียก /iáók/ [154-clam.wav] ‘seashell’ ปู้เตียก /putiák/ [155-white.wav] ‘white’ ชู้วียก /c <sup>h</sup> uwiák/ [156-rod.wav] ‘large fish hook’	
/uə/	<๓>	เกือะ /kuəʔ/ [157-shoe.wav] ‘shoe’ บี่ลื้อย /bəlúəj/ [036-weary.wav] ‘achy’	
/uə/	<๓>	ลัว /luá/ [158-elder brother’s wife.wav] ‘elder brother’s wife’ แปจัวะ /pɛcuáʔ/ [159-hoe.wav] ‘hoe’	In open syllables or syllables ending with /-ʔ/ or /-h/
	<๓>	บี่ดาวก /bituák/ [160-star.wav] ‘star’ ทวด /t <sup>h</sup> uát/ ‘great-grandparent’	In closed-syllables not ending with /-ʔ/ or /-h/
/ə/	<๓>	ดี่ไล /dəláj/ [162-mountain.wav] ‘mountain’ ดี่เตย /təláj/ [163-three.wav] ‘three’ กั๊ง /kədín/ [164-bell.wav] ‘small bell’	

Two important points should be noted. First, the vowels /a/, /i/, /u/, /ɛ/, and /ɔ/ in unstressed positions are uniformly written with symbols for long vowels <ᳵ>, <ᳶ>, <᳷>, <᳸>, and <᳹> rather than their short counterparts <ᳵ̄>, <ᳶ̄>, <᳷̄>, <᳸̄>, and <᳹̄>. This is because, as discussed in §3, the unstressed vowels in Moklen are relatively long. Moreover, in Thai orthography, symbols for short vowels are associated with the glottal stop so that they are often read as /aʔ/, /iʔ/, /uʔ/, /ɛʔ/, and /ɔʔ/ (Haas 1956: 17). Second, the neutral vowel /ə/ is always written as <ᳶ> instead of the symbol <ᳵ̄>, which Kraissame (2016: 61) proposes for the corresponding segment in Moken. This is also because <ᳵ̄> is strongly associated with a final glottal stop /ʔ/, most often pronounced as /aʔ/.

In addition to consonants and vowels, the Moklen orthography also needs symbols to signal the tonal distinction. Given that an overwhelming majority of lexical items have Tone 1, most words in Moklen can be left unmarked, and words with Tone 2 can be considered as special cases. For example, /kɔ̀lɑ̀:t/ [039-hot.wav] ‘hot’ and /kɔ̀lɑ̀:t/ [040-mushroom.wav] ‘mushroom’ are written as กอลาท and กอลาท<sup>๑</sup>, respectively. Among the five tone marks in the Thai orthography, <ᳶ> is selected for maximum ease of learning because it is the least frequent and can only be interpreted as a rising tone in Standard Thai. This choice avoids confusion that may arise due to the many-to-many correspondences between Thai lexical tones and tone marks. The two most common tone marks <ᳵ> and <᳷> are not suitable because their phonemic values fluctuate depending on whether the initial consonant letters belong to the low, mid, or high series. Moreover, <᳷̄> is not chosen because it represents Tone 4 in Standard Thai, which is often realized phonetically closer to Moklen Tone 1. Note that the shortener <ᳶ̄> must be dropped in words marked with <ᳶ>, as Thai orthography does not allow both on top of each other (Haas 1956: 73–76). For example, /bɔ̀j/ [171-get.wav] ‘to get’ is written as เบ็๑, but /bɔ̀j/ [169-2SG.wav] ‘2s pron.’ is spelled without <ᳶ̄> as เบ็.

Lastly, the proposed orthography does not use a space to signal word boundaries but instead uses it between sentences. Although Premsrirat (2008) recommends putting spaces between words in Thai-based orthography for minority languages, the practice would lead to complications in writing Moklen. First, attempting to decide between compounds and phrases would be difficult both for community members and outsiders. For example, it would not be clear whether ‘fish curry’ [166-fish curry.wav] should be written as ชอนบช แอกรน or ชอนบชแอกรน. The former assumes that ‘fish’ is a phrasal modifier that specifies the main ingredient of the curry, while the latter views the entire expression as a name for a specific type of curry. In addition, putting spaces between words may confuse community members who are used to the Thai convention.

## 6. Discussion

The research team worked closely with the community leaders in gaining constant feedback from various groups of speakers since the beginning of the design. However, a formal evaluation of the orthography by the community at large is still needed in order to validate the usability and acceptance of the proposed orthography. While an intensive investigation into the acceptance of the orthography is needed for a further analysis, certain potential factors that may affect the adoption of the orthography among different Moklen communities can be anticipated based on our observations during the data collection.

We relied on the data collected from two Moklen communities, due to existing social relationships and convenience, to inform our design decisions. Nonetheless, there are almost twenty Moklen communities located in other parts of Phang Nga and its neighboring provinces (Arunotai 2017). Our goal is to develop an orthography that can be used by all Moklen community members, as well as outsiders who would like to learn Moklen. Therefore, it is important to ensure its acceptance by members of the Moklen community at large. This can be done through a follow-up evaluation with an appropriate sampling technique. Nevertheless, it is important to note that, to date, there are no systematic surveys of the language situation that records the number of people who speak the language on a daily basis, their level of fluency, their demographic characteristics, and their attitudes toward the use of language. The lack of a complete landscape of the Moklen speakers is definitely one of the challenges for a systematic sampling approach.

It is also evident that age is one of the major determinants in speaking Moklen. The majority of our language consultants were elderly and were available at their residences during the day, while the working-age population was busy making a living. Although the current community leaders were of the working-age group, they expressed concerns about the effects that exposure to and adoption of Thai social and cultural lifestyles from a very young age may have on the vitality of their language. Moreover, as mentioned above, we were able to interact with school-aged community

members and found that some of them associated speaking Moklen as being alienating and disadvantageous, while some were explicitly eager to learn and use the language. An in-depth investigation on perceptions of community members from various generations would contribute to a provision of language adoption among incoming and future community members, which is essentially linked to the survival of the language.

The sociopolitical identities of Moklen and related minority ethnic groups are becoming more hybrid and complex (Robinson & Drozdowski 2016). While a number of Moklen who sought economic and social opportunities have left the communities and, therefore, may have abandoned the language for various reasons, another group of community members who may play an important role in learning and using the language are adopted/inducted community members (e.g., via intermarriage or migration). Robinson & Drozdowski (2016: 548) particularly see intermarriage as the biggest barrier in language and cultural transfer. Their support to the descendants to learn and use the Moklen language, as well as to document community knowledge, is essential to the vitality of the language and the identity of the Moklen as well. Therefore, their perception of the usage of Moklen may also be worthy of further investigation.

Designing an orthography is a part of an effort to preserve the language. However, in a community where the use of the language is declining, language revitalization should also be taken into consideration. The revitalization effort should involve more than the use of the language within the household. Instead, it should be accepted in different parts of communities (e.g., businesses, schools, and government offices). As a minority population group, community leaders at Bangsak think that language preservation and revitalization are more feasible via a formal education system and thus emphasize creating pedagogical materials. A similar approach has been implemented with other minority languages in Thailand, such as Pattani Malay (Premsrirat & Samoh 2012).

Because the proposed Thai-based orthography is intended to facilitate the Thai-literate sector of the Moklen population, the few discrepancies between Moklen and Thai phonology may cause confusion in the initial stage of learning. Effective pedagogical approaches to minimize cognitive load, confusion, and learning effort are still to be explored (Taylor-Adams 2019). This is one of the major challenges in preserving the language in writing, particularly when there are very few native-speaking teachers and very few learning resources. In addition, since the adoption rate of digital technology among Moklen community members has been continuously increasing across all generations, then the various modes, methods, and media of digital technology can play essential roles in the dissemination and promotion of the Moklen language, which have already been implemented in the documentation and preservation of other indigenous languages (e.g., Lhawa 2019; Chew 2021).

As a part of language and cultural heritage preservation, the community leaders also wish to increase the number of written language-learning resources to facilitate informal learning environments. These resources would also help community members who are not in school to learn the language as well. As an initial effort to develop the written language, we are also developing a Moklen–Thai–English glossary, using the orthography proposed here for both lay users and linguists. We hope to engage Moklen community members in expanding the lexical database and to assist them in recording and writing their language as well as their intangible cultural heritage.

## 7. Conclusion

This paper addresses the initial phase of Moklen's orthographic design based on data collected during several field trips to Bangsak Moklen communities. Adopting UNESCO's framework (Robinson & Gadelii 2003), we advocate for the use of a Thai-based orthography to write the Moklen language, having considered all the key factors, including social relations, government support, cultural heritage, and language structure. Moreover, we carefully considered Smalley's (1959) adequacy criteria in making decisions about specific orthography design problems. The result is a Thai-based orthography of Moklen that not only captures all the phonemic distinctions but could also facilitate learning, boost literacy, encourage production of reading and pedagogical materials, and, most importantly, inspire members of the Moklen community to write in their own language. However, the development of orthography is only the first step in a rush to produce a written documentation of the Moklen language, traditional knowledge, and oral history, in the face of its alarming state of endangerment.

## Appendix

This “Crow and Jackal” text was collected by Daniel Loss from a fifty-year-old female Moklen speaker from Baan Thung Waa, Khuek Khak, Takua Pa, Phang Nga, using a picture-task story designed for use by the Social Cognition and Language Project (Carroll et al. 2011). Instances of code-switching are underlined, and the Standard Thai transcription for the words is provided in brackets.<sup>6</sup>

- 1      *ตีจุมนี่ตุ๋น*  
          ticúm            nidú:n  
          bird            sleep  
          ‘A bird is sleeping.’
- 2      *ตีจุมนี่ตุ๋น*  
          ticúm            nidú:n  
          bird            sleep  
          ‘The bird is sleeping.’
- 3      *ตีจุมจู้ว่แอกาน*  
          ticúm            nám            ʔeká:n  
          bird            eat            fish  
          ‘The bird is eating a fish.’
- 4      *เนยเฮีย(เฮีย)*  
          ...            nɔ̀j            ʔɔ̀j            (ʔɔ̀j)  
          HES            this            dog            SFP  
          ‘Here’s a dog.’
- 5      *ตีจุมบนำแอกานเอ็กคากาเอา*  
          ticúm            bá:ʔ            ʔeká:n            ʔék            tá:            kaʔé:w  
          bird            carry            fish            ascend            on            wood  
          ‘The bird takes a fish and goes up a tree.’
- 6      *ตีจุมแอมคากาเอา*            *แอกานแอมเสห้ออกิ่ง*  
          ticúm            ʔé:m            datá:            kaʔé:w            ʔeká:n            ʔé:m            héʔ            ʔokáj  
          bird            exist            on            wood            fish            exist            at            mouth  
          ‘The bird is on the tree. The fish is in its mouth.’

<sup>6</sup> HES = hesitation; SFP = sentence-final particle

- 7      *เนยเี้ยนาเตียง*  
 nɤ̀j            ʔɤ̀j            natióŋ  
 this            dog            wait  
 ‘Here, the dog’s waiting.’
- 8      *เี้ย                    เอนยอนยอแยแล้วนาเตียงแอกาน*  
 ʔɤ̀j            (.)            ʔe:n-jón            [jɔ̀jʔ]            [lɛ:wʔ]            natióŋ            ʔeká:n  
 dog                            water-saliva            drip            already            wait            fish  
 ‘The dog’s saliva is dripping, waiting for the fish.’
- 9      *ตีจุม                    บูลัดเนยคูแอกานแอมคาตากาแาว*  
 ticúm            (.)            bulát            nɤ̀j            kʰú:            ʔeká:n            ʔé:m            datá:            kaʔé:w  
 bird                            CLF            this            with/and            fish            exist            on            wood  
 ‘This bird here and the fish are on the tree.’
- 10     *อาลัดเนยแมตาท*  
 ʔa=lát            nɤ̀j            metá:ʔ  
 one=CLF            this            request  
 ‘This one, it’s requesting.’
- 11     *ออแก้งเปิด                    แมตาทแอกานตั้งลัดเนยลัดเนยลัดเนยลัดเนย*  
 ʔokáŋ            [pɤ̀:t]            metá:ʔ            ʔeká:n            táŋ            lát            nɤ̀j            (lát nɤ̀j lát nɤ̀j)  
 mouth            open            request            fish            from            CLF            this            SFP  
 ‘Its mouth is open. It’s requesting the fish from it, this one.’
- 12     *เี้ยเนยนาเตียงอ้วปลัดญาเบี่ยอู้า*  
 ʔɤ̀j            nɤ̀j            natióŋ            [tʰa:]            pʰlà:t            ná:            bɤ̀j            nám  
 dog            this            wait            if            fall            3SG            be.able            eat  
 ‘The dog is waiting. If it falls, it’ll get to eat it.’
- 13     *ตีจุมเนยเม้นอู้าเบี่ยชะฮ*  
 ticúm            nɤ̀j            mén                            nám                            bɤ̀j                            háh  
 bird            this            take                            eat                            be.able                            NEG  
 ‘The bird won’t eat it.’

- 14 *ตีจุ่มนยปล่อยแอกานดินพลัด้อนเอี้ยจู้*  
 ticúm nɔ̀j [plɔ̀:j] ʔɛká:n dín pʰlà:t ʔón ʔɔ̀j nám  
 bird here release fish come fall give dog eat  
 ‘The bird releases the fish, so it falls and the dog can eat it.’
- 15 *อ้อนเอี้ยออกังงานเตียง*  
 ʔón ʔɔ̀j ʔokáŋ natiəŋ  
 give dog mouth wait  
 ‘It gives it to the dog. Its mouth is waiting.’
- 16 *เอี้ยเบี้ยจู้แอกานแล้ว*  
 ʔɔ̀j bɔ̀j nám ʔɛká:n [lɛ:wɿ]  
 dog acquire eat fish already  
 ‘The dog has gotten to eat the fish!’
- 17 *บี้ดิ่งนิตูนสบาย*  
 bətáŋ nidú:n [tɛʰəbá:j-]  
 be.full sleep be.relaxing  
 ‘It’s full. It’s sleeping comfortably.’

## References

- Arunotai, Narumon. 2017. “Hopeless at sea, landless on shore”: Contextualizing the sea nomads’ dilemma in Thailand. *AAS Working Papers in Social Anthropology* 31. 1–27. [doi:10.1553/wpsa31](https://doi.org/10.1553/wpsa31)
- Blust, Robert. 1994. The Austronesian settlement of Mainland Southeast Asia. In Adams, Karen L. & Thomas J. Hudak (eds.), *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, 25–83. Tempe, AZ: Arizona State University Program for Southeast Asian Studies.
- Cahill, Michael. 2014. Non-linguistic factors in orthographies. In Cahill, Michael & Keren Rice (eds.), *Developing orthographies for unwritten languages*, 9–25. Dallas, TX: SIL International.
- Carroll, Alice, Barbara Kelly, & Lauren Gawne. 2011. The jackal and crow picture task. Designed for use by the Social Cognition and Language Project. A collaboration of the Australian National University, Griffith University, University of Melbourne, and the Max Planck Institute for Psycholinguistics. (<http://tulquest.huma-num.fr/en/node/161>) (Accessed 2023-03-29.)
- Chew, Kari A. B. 2021. #KeepOurLanguagesStrong: Indigenous language revitalization on social media during the early COVID-19 pandemic. *Language Documentation & Conservation* 15. 239–266. (<http://hdl.handle.net/10125/24976>)



- Cooke, Joseph R., J. Edwin Hudspith, & James A. Morris. 1976. Phlong (Pwo Karen of Hot District, Chiang Mai). In Smalley, William A. (ed.), *Phonemes and orthography: Language planning in ten minority languages of Thailand*, 187–220. Canberra: Pacific Linguistics.
- Coulmas, Florian. 2003. *Writing systems: An introduction to their linguistic analysis*. Cambridge, UK: Cambridge University Press.
- Danvivathana, Nantana. 1987. *The Thai writing system*. Hamburg, Germany: Helmut Buske verlag.
- Filbeck, David. 1976. Mal (Thin). In Smalley, William A. (ed.), *Phonemes and orthography: Language planning in ten minority languages of Thailand*, 239–257. Canberra, Australia: Pacific Linguistics.
- Grenoble, Lenore A. & Lindsay J. Whaley. 2006. *Saving languages: An introduction to language revitalization*. Cambridge, UK: Cambridge University Press.
- Haas, Mary R. 1956. *The Thai system of writing*. Washington, DC: American Council of Learned Societies.
- Hogan, David W. 1972. Men of the sea: Coastal tribes of South Thailand's west coast. *Journal of the Siam Society* 60. 205–235.
- Hogan, David W. 1976. Urak Lawoi' (Orang Laut). In Smalley, William A. (ed.), *Phonemes and orthography: Language planning in ten minority languages of Thailand*, 283–302. Canberra: Pacific Linguistics.
- Kraisame, Sarawut. 2016. Process of documentation and preservation of an endangered language: A case study of Moken on Surin Islands. Nakhon Pathom, Thailand: Mahidol University. (Unpublished doctoral dissertation).
- Kraisame, Sarawut. 2018. Language endangerment and community empowerment: Experience from community training in the Moken language documentation and preservation project. *Kasetsart Journal of Social Sciences* 39(2). 244–253. [doi:10.1016/j.kjss.2017.05.002](https://doi.org/10.1016/j.kjss.2017.05.002)
- Larish, Michael D. 1992. Who are the Moken and Moklen on the islands and coasts of the Andaman Sea? *Pan-Asiatic linguistics: Proceedings of the Third International Symposium on Language and Linguistics*, Bangkok, 8–10 January, vol. 3, 1305–1319. Bangkok: Chulalongkorn University Printing House.
- Larish, Michael D. 1997. Moklen-Moken phonology: Mainland or insular Southeast Asian typology? In Odé, Cecilia & Wim A. L. Stokhof (eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 125–150. Amsterdam: Rodopi.
- Larish, Michael D. 1999. *The position of Moken and Moklen within the Austronesian language family*. Honolulu, HI: University of Hawai'i Press.
- Larish, Michael D. 2005. Moken and Moklen. In Adelaar, Alexander & Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 513–533. London: Routledge.
- Lhawa, Yulha. 2019. Language revitalization, video, and mobile social media: A case study from the Khroskyabs language amongst Tibetans in China. *Language Documentation & Conservation* 13. 564–579. (<http://hdl.handle.net/10125/24902>)
- Lüpke, Friederike. 2011. Orthography development. In Austin, Peter K. & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 312–336. Cambridge, UK: Cambridge University Press. [doi:10.1017/CBO9780511975981.016](https://doi.org/10.1017/CBO9780511975981.016)
- Office of the Royal Society. 2018. Strategic plan for National Language Policy, 2018–2021 [Draft]. (<http://legacy.orst.go.th/wp-content/uploads/2018/11/%E0%B8%A3%E0%B9%88%E0%B8%B2%E0%B8%87-%E0%B9%81%E0%B8%9C%E0%B8%99%E0%B8%A2%E0%B8%B8%E0%B8%97%E0%B8%98%E0%B8%A8%E0%B8%B2%E0%B8%AA%E0%B8%95%E0%B8%A3%E0%B9%8C%E0%B8%99%E0%B9%82%E0%B8%A2%E0%B8%9A%E0%B8%B2%E0%B8%A2%E0%B8%A0%E0%B8%B2%E0%B8%A9%E0%B8%B2%E0%B8%8A%E0%B8%B2%E0%B8%95%E0%B8%B4-%E0%B8%AA%E0%B9%88%E0%B8%A7%E0%B8%991.pdf> 6) (Accessed 2024-01-11.)
- Person, Kirk R. 2001. Writing Bisu: A community-based approach to orthography development. In Thurgood, Graham W. (ed.), *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, 171–200. Tempe, AZ: Arizona State University, Program for Southeast Asian Studies.

- Pittayaporn, Pittayawat. 2005. Moken as a Mainland Southeast Asian language. In Grant, Anthony & Paul Sidwell (eds.), *Chamic and beyond*, 189–210. Canberra: Pacific Linguistics.
- Pittayaporn, Pittayawat, Warunsiri Pornpottanamas, & Daniel Loss (eds.). 2022. *Moklen-Thai-English Dictionary: Pilot version*. Bangkok: Academic Work Dissemination Project, Faculty of Arts, Chulalongkorn University.
- Premrsirat, Suwilai. 2000. Developing Chong writing system. *Journal of Language and Culture* 19(2). 5–18.
- Premrsirat, Suwilai. 2008. Orthography development: A tool for revitalizing and maintaining ethnic minority languages. *Journal of Language and Culture* 26. 18–34.
- Premrsirat, Suwilai. 2018a. Mahidol model for the preservation of language diversity: Thailand experience. In Premrsirat, Suwilai & David Hirsh (eds.), *Language revitalization: Insights from Thailand*, 27–44. Bern: Peter Lang.
- Premrsirat, Suwilai. 2018b. Orthography development: A tool for language revitalization and preservation of local wisdom. In Premrsirat, Suwilai & David Hirsh (eds.), *Language revitalization: Insights from Thailand*, 45–77. Bern: Peter Lang.
- Premrsirat, Suwilai & Uniansasmita Samoh. 2012. Planning and implementing Patani Malay in bilingual education in southern Thailand. *Journal of the Southeast Asian Linguistics Society* 5. 85–96.
- Robinson, Clinton & Karl Gadelii. 2003. *Writing unwritten languages: A guide to the process; working paper*. UNESCO.
- Robinson, Daniel F. & Danielle Drozdowski. 2016. Hybrid identities: Juxtaposing multiple identities against the ‘authentic’ Moken. *Identities* 23(5). 536–554. [doi:10.1080/1070289X.2015.1070730](https://doi.org/10.1080/1070289X.2015.1070730)
- Taylor-Adams, Allison. 2019. Recording to revitalize: Language teachers and documentation design. *Language Documentation & Conservation* 13. 426–445. (<http://hdl.handle.net/10125/24873>)
- The Royal Society of Thailand. 2010. *Manual of the Thai-based Patani Malay orthography*. Bangkok: Office of the Royal Society.
- The Royal Society of Thailand. 2018. *Manual of the Thai-based Gong orthography*. Bangkok: Office of the Royal Society.
- Samoh, Uniansasmita. 2018. Languages and scripts reflecting Patani Malay multiple identities in Thailand’s Deep South. *Journal of the Southeast Asian Linguistics Society* 11(2). cxi–cxxiv. (<http://hdl.handle.net/10524/52439>)
- Seifart, Frank. 2006. Orthography development. In Gippert, Jost, Nikolaus P. Himmelmann, & Ulrike Mosel (eds.), *Essentials of language documentation*, 275–299. Berlin: Mouton de Gruyter.
- Smalley, William A. 1959. How shall I write this language? *The Bible Translator* 10(2). 49–69. [doi:10.1177/000608445901000201](https://doi.org/10.1177/000608445901000201)
- Smalley, William A. 1976. Writing systems in Thailand’s marginal languages: History and policy. In Smalley, William A. (ed.), *Phonemes and orthography: Language planning in ten minority languages of Thailand*, 1–24. Canberra: Pacific Linguistics.
- Swastham, Pensiri. 1982. *A description of Moklen: A Malayo-Polynesian language*. Nakhon Pathom, Thailand: Mahidol University.
- UNESCO Ad Hoc Expert Group on Endangered Languages. 2003. *Language vitality and endangerment*. (<https://ich.unesco.org/doc/src/00120-EN.pdf>) (Accessed 2024-01-19.)
- Ungpho, Rewadee. 2016. Communication through melody and lyrics: Cultural identity of the Moklen. *Thammasat Review* 19(2). 22–41.
- United Nations. 2007. United Nations Declaration on the Rights of Indigenous Peoples. New York, NY: United Nations. (<https://social.desa.un.org/issues/indigenous-peoples/united-nations-declaration-on-the-rights-of-indigenous-peoples>) (Accessed 2024-01-11.)
- Willis Oko, Christina M. 2018. Orthography development for Darma (The case that wasn’t). *Language Documentation & Conservation* 12(1). 15–46. (<http://hdl.handle.net/10125/24761>)