

Generic combining marks in Thai script minority languages

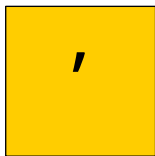
Ben Mitchell

6 May 2026

This document concerns two modifier signs and five generic combining marks, and in particular their use in minority language orthographies that use the Thai script. We suggest Unicode updates the official recommendations in Chapter 16 of the Standard to include the notes marked with 🖱️. We also propose the encoding of a new character, THAI GLOTTAL STOP.

Martin Hosken's *Proposal not to encode 4 minority Thai letters for Patani Malay* (2010) details the use of MODIFIER LETTER APOSTROPHE (u+02BC), MODIFIER LETTER MINUS (u+02D7), COMBINING TILDE (u+0303) and COMBINING MACRON BELOW (u+0331); these characters are now specified in some detail in Chapter 16 of the Unicode standard for Patani Malay language in Thai script. These characters are used in other languages as follows:

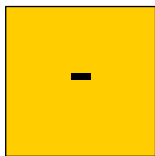
1. MODIFIER LETTER APOSTROPHE (u+02BC)



As Martin explains, this is used¹ in the official Thai-script orthography for Patani Malay language, to mark glottalisation. It is stored and appears directly before a consonant letter, e.g. 'ก, 'ก, 'ง, 'ช, 'ด, 'ต, 'บ, 'ล, 'ย and 'น.

🖱️ Western Lawa orthography uses u+02BC in the same way to mark glottalisation, e.g. 'ง, 'ญ, 'น, 'ม, 'ย, 'ร and 'ล. In sequences with a prebase vowel, e.g. โ้ระ we can see the glottalisation mark stays directly before the consonant, with the prebase vowel to its left. Kerning improves the appearance for some sequences.

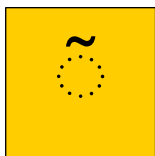
2. MODIFIER LETTER MINUS (u+02D7)



Patani Malay uses this character to mark compound vowels (diphthongs) that do not occur in Thai language, e.g. ာ-ဝီ, ာ-ဝီ, ာ-ဝူ, ာ-ဝူ.

🖱️ We also find the same usage in Western Lawa (L'wa) and Yoy orthographies, e.g. ာ-ဝီ, ູ-ແ, ເ-ဝီ. Some users type a hyphen instead, but this would allow vowel sequences to linebreak and would be best avoided.

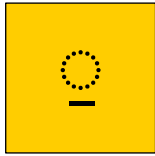
3. COMBINING TILDE (u+0303)



Combining tilde is used only in Patani Malay, where it sits above base consonants to mark nasalisation in certain vowel sequences, e.g. ൣ်း, ൣ်း, ൣ်း, ൣ်း, ൣ်း, ൣ်း. This presents no problems with normalisation as the tilde would be stored and displayed at the end of a cluster, following any other diacritics.

¹ Mahidol University's documentation refers to this glottalisation mark as FON THONG (ฟนทอง). Fon Thong, cognate with KHMER SIGN BANTOC (u+17CB), is an as-yet unencoded Thai character found in historic manuscripts, where it represents a short vowel sound (not the consonant modifier we need here). Additionally fon thong is a combining mark sitting above a consonant, rather than a spacing character to the left of a consonant. Therefore, both semantically and behaviourally, using fon thong to mark glottalisation would be inappropriate.

4. COMBINING MACRON BELOW (u+0331)



Combining macron below is specified for Patani Malay, where it marks combined consonants with a nasal aspect, e.g. **๓** /ŋ^g/, **๔** /ŋ^j/, **๕** /n^d/, and **๖** /m^b/.

☞ Similarly, Western Lawa marks consonants with macronbelow, e.g. **๗** /^g/, **๘** /^j/, **๙** /^d/ and **๑๐** /^b/.

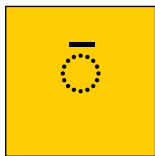
☞ For Northern Pwo, the Non-Formal Education Department (NFE) orthography (2003) uses the macronbelow to mark consonant letters whose sound value differs from Standard Thai, e.g. **๑๑** /x/, **๑๒** /ɰ/ and **๑๓** /ŋ/.

☞ On the other hand, Moken language uses macronbelow with vowel letters to mark vowels that do not occur in Thai language, e.g. **๑๔**, **๑๕**.

As Martin notes, normalisation places macronbelow after any Thai diacritics, so fonts should reorder it to display in the correct position, directly after the preceding base letter before any other diacritics.

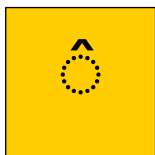
Following the specifications above, we suggest the following generic combining marks for other orthographies and suggest Unicode includes an official recommendation:

5. COMBINING MACRON (u+0304)



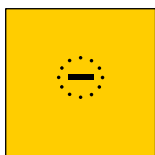
☞ Combining macron appears in Saek orthography, where it is used as a tonemark. This means it can sit on a base consonant directly, e.g. **๑๖**, **๑๗**, **๑๘**, or if the consonant also carries a vowel diacritic above, the macron will shift upwards to sit on the vowel, e.g. **๑๙**, **๒๐**. This presents no problems with normalisation, as tonemarks are stored and displayed at the end of a cluster, following any other diacritics.

6. COMBINING CIRCUMFLEX ACCENT (u+0302)



☞ Saek language also uses circumflex as a tonemark. Like combining macron, this sits on a base consonant directly, e.g. **๒๑**, **๒๒**, **๒๓**, or if the consonant also carries a vowel diacritic above, the circumflex will shift upwards to sit on the vowel, e.g. **๒๔**, **๒๕**. This presents no problems with normalisation, as tonemarks are stored and displayed at the end of a cluster, following any other diacritics.

7. COMBINING SHORT STROKE OVERLAY (u+0335)



☞ Mpi language uses the stroke overlay with the vowel letter SARA AA (u+0E32) to represent a vowel that does not occur in Standard Thai, i.e. **๒๖**. This presents no problems with normalisation, as saraAa does not carry any other diacritics in Mpi language.

Finally we propose the encoding of one character for Nyah Kur language:

8. THAI GLOTTAL STOP (u+0E3B?)



Nyah Kur language adds one extra base letter to indicate a glottal stop following a vowel. This character seems to occur only in syllable-final position, where it cannot carry any diacritic marks. As this character is not currently encoded, users are resorting to using the Latin question mark, or the Latin glottal stop letter ʔ (u+0294). The question mark is clearly not advisable as it's very likely to be interpreted as marking a question. The Latin glottal stop is not the best solution as we should avoid breaking script runs within a single word (to allow kerning, for example), and it is preferable for this character to have a Thai-specific size and design. We therefore propose encoding a Thai glottal stop separately.

As a point of reference, the Devanagari range includes a script-specific glottal stop at u+097D.

Character properties

```
0E01;THAI CHARACTER GLOTTAL STOP;Lo;0;L;;;;;N;;;;;
```

Attestations of Thai glottal stop

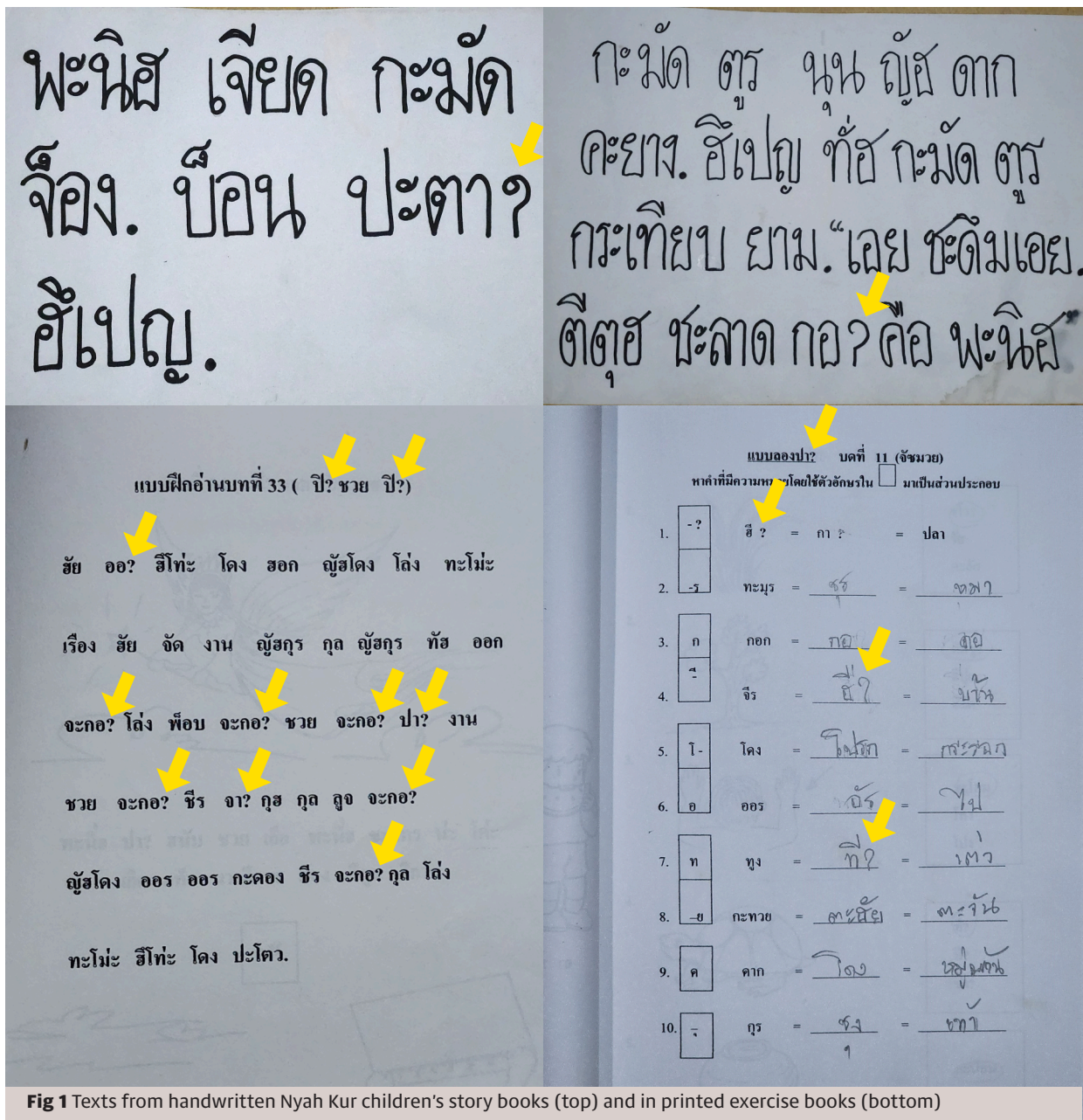


Fig 1 Texts from handwritten Nyah Kur children's story books (top) and in printed exercise books (bottom)

ในภาษาญักรจะใช้เครื่องหมาย ? แทนพยัญชนะกักที่ตามหลังเสียงสระยาวทำให้
สระเสียงยาวออกเสียงสั้นลงแต่ไม่สั้นเท่ากับสระเสียงสั้นในภาษาไทย ดังตัวอย่างเพิ่มเติม
ต่อไปนี้

- ตี? = ดิน
- ที? = เต่า
- ฮึ? = บ้าน
- ปา? = ทำ
- จา? = กิน
- ฮนา? = กระแต
- ฮนึ? = หนู
- โซร? = ข้าวเปลือก

ตัวอย่างเรื่องสั้นและนิทาน

ฮึโด?
หอย

นอม เทวดา มวย อ้อ นัย อ้อ กะขอ? กุล พะนิฮ. กุล จา? ไป ปี้? ฮัย มวย ต้อง.
มี เทวดา ๑ องค์ จะ ไป บอก ให้ คน ให้ กิน ข้าว ๓ วัน ๑ ครั้ง

เทวดา อ้อ พ็อบ ชุร. กุล ชุร อ้อ กะขอ? กุล พะนิฮ.
เทวดา ไป พบ, เจอ หมา ให้ หมา ไป บอก ให้ คน

ชุร พ็อบ เปรียง. กะขอ? กุล เปรียง อ้อ กะขอ? กุล พะนิฮ.
หมา พบ ควาย บอก ให้ ควาย ไป บอก ให้ คน

เปรียง อ้อ กะขอ? ลูจ. กุล จา? ไป มวย ฮัย ปี้? ต้อง.
ควาย ไป บอก ผิด ให้ กิน ข้าว ๑ วัน ๓ ครั้ง

เทวดา ฮึร กุล ชุร กุฮ อ้อ กะขอ? เณญ.
เทวดา โมโห ให้ หมา ไม่ ไป บอก เอง

เลย กุล ชุร จา? อิจ.
เลย ให้ หมา กิน อัจจาระ

เปรียง ตี แต้ฮ ทีบ กุล แต้ฮ จา? ฮึพัด.
ควาย ที่ มัน ไร่ ให้ มัน กิน หล้า

เทวดา กอ อ้อ กะขอ? เณญ.

Fig 2 Pages from Nyah Kur textbook using the Latin glottal stop character



Fig 3 Instagram post showing glottal stop character in video subtitles

๓.๒ ประเพณีแห่ทอดอกผึ้ง

๑) เรื่อง นั้งเดิม นั้งด้อฮ คอง ฮอ กาวพลุย

แท้ฮ ปะเพนี คอง เปญยอง ปา? ฮัย ตรุด แท้ฮ งาน บุน อะโต คอง พะนีฮ ญ้อกร. นะ เริ่ม ปา? ฮัย ที ฮีจิด คอง ฮีตู? ซุน. ญ้อโดง นะ เตรียมงาน ซีรคอง ตรุยตรุย นะ อ้อ ซีร ชู? โลง ปา? โครง ฮอ กาว พลุย เพราะเพราะ นะปา? คองจา? เลียง พะนีฮ โลง ปา? งาน. ฮัยที จัซ กอ นะ ปา? พักชะกอ? นอม ญ้อฮ ปะชู่ คอย ปะโตว ตี ปา? ญั้งแอฮญั้งแตะ. ฮัยที จัซมวย จัซบาร นะ ซวย ชะกอ? ปา? กุล ทูล. พอ ฮัยที จัซปี? ซวย ชะกอ? กะเดียน กาวพลุย กุล ทูล ฮัยค้อฮ. ปะตัม กอ นะ วิญ ชะกอ? ปา? เรเร กุล ชะนุก ชะนาน. พอ ตอน นั้ง แรฮ ญ้อโดง จา?โปง ทูล กอ นะ ติด อ้อ บอน ซาลา ดุกญ้อฮ. พอ แล้จ เวลา กอ นะ พา ชะกอ แฮ ฮอ กาวพลุย (พาปา ค่อ) ร่อบ โดง เท็ก ทะโพน รัม อ้อ แล้จ วัต แฮ ร่อบ วัต ปี? ร่อบ ทูล กอ เจียด อ้อ ทวาย ลุง รับ ซิน รับ พอน กอ ทูล.

Fig 4 Digital text in Nyah Kur language, using question mark in place of glottal stop

References

- **ธีรภาพ เขื่อนสี และคณะ** (2010), *สร้างระบบตัวเขียน: สิบชะตาภาษาอิมบ้านดงตำบลสวนเขื่อน อำเภอเมือง จังหวัดแพร่*
- **นิล กล้าทะเล และคณะ** (2016), *ลาญาน มอแกน (เรื่องราว ชาวมอแกน)*
- **ปัญญา บริสุทธิ์** (2010), *คู่มือระบบเขียนภาษามลายูปาตานิ อักษรไทย*
- **เปรมศรีรัตน์, สุวิไล, มยุรี การพัฒนา และ กุมารี ลากอาภรณ์** (2013?), *โครงการบันทึกรวบรวมเพื่อการสงวนรักษา ภาษาและภูมิปัญญาท้องถิ่นของกลุ่มชาติพันธุ์: ภาษาญ้อกูร*
- **ประพันธ์ จันทร์ยวง และคณะ** (2022), *เลขขอมแล เลขเวีอะ*
- **วัฒนธรรม ปีที่ 54 ฉบับที่ 2 เมษายน-มิถุนายน** (2015)
- **กองศิลปกรรม** (2015), *คู่มือระบบเขียนภาษาเลขเวีอะอักษรไทย ฉบับราชบัณฑิตยสภา*
- **สำนักงานราชบัณฑิตยสภา** (2015), *คู่มือระบบเขียนภาษาญ้อกูรอักษรไทย ฉบับราชบัณฑิตยสภา*
- **กรมส่งเสริมวัฒนธรรม** (2015), *ภาษา: มรดกภูมิปัญญาทางวัฒนธรรมของชาติ*
- **ปรีชา ชัยปัญหา และคณะ** (2010), *แนวทางการจัดทำพจนานุกรมภาษาแสกฉบับชาวบ้านเพื่อการอนุรักษ์และฟื้นฟูภาษาและภูมิปัญญาแสกบ้านบะหว้า ต.ท่าเรือ อ.นาหว้า จ.นครพนม. รายงานวิจัยสำนักงานกองทุนสนับสนุนการวิจัย (สกว.)*
- **วิจัยไ้ยย** (undated), *ระบบเขียนภาษาไ้ยยพะยะญุชณะต้นประกอบด้วย 19 ตัว*
- **Constable, Peter** (2018), *Canonical Ordering of Marks in Thai Script*
- **Hosken, Martin** (2010), *Proposal not to encode 4 minority Thai letters for Patani Malay*
- **Phillips, Audra** (2014), *Omkoi Pwo Karen Phonology and Orthography*
- **Smalley, William A (ed)** (1976), *Phonemes and Orthography: Language planning in ten minority languages of Thailand*
- **Tatsuro Daniel Arisawa** (2016), *An Iu Mien Grammar: a tool for language documentation and revitalisation*
- **TU-SIL-LRDP Committee** (2002), *Minority Language Orthography in Thailand: five case studies*