

CHRONOLOGY OF REGISTROGENESIS IN KHMER: ANALYSES OF POETRY AND INSCRIPTIONS¹

Sireemas MASPONG

Institute for Phonetics and Speech Processing (IPS), LMU Munich
s.maspong@phonetik.uni-muenchen.de

Abstract

This study examines the phonological changes in Khmer, focusing on the loss of onset voicing, the emergence of register contrast, and the development of a bifurcated vowel system. By analyzing rhyme pairs from Khmer poetry (16th-19th centuries) and spelling variations in Khmer inscriptions (late 7th-mid-18th centuries), I propose a chronology of these changes. The results indicate that vowel mergers resulting from vowel bifurcation in the rhyme pairs appeared from the 17th century onwards. Additionally, spelling variations reflecting the loss of onset voicing and the vowel mergers were observed in inscriptions dating back to the late 16th century. These findings align with previous research, suggesting that both the loss of onset voicing and the vowel bifurcation occurred in the late 16th century. However, the relative chronology of these changes remains uncertain since evidence for both changes emerges around the same time.

Keywords: Khmer, registrogenesis, chronology
ISO 639-3 codes: khm

1 Introduction

Khmer is one of few Austroasiatic languages with a long history of written documents dating back to the 7th century. It has therefore been very important for the understanding of registrogenesis. In comparison to other languages in the Austroasiatic family, Khmer is considered an innovative language (Huffman 1976: 58–59) because of the extensive phonological restructuring that took place, especially in the Middle Khmer period (14th-18th century). Reconstructions of proto-Khmer (Ferlus 1992:81) and Khmer orthography from the 7th century until the present day indicate that the language had a voicing contrast onset position. Standard Khmer, a contemporary descendent language, lost this contrast as a consequence of the devoicing of voiced stops. The loss of voicing contrast resulted in a merger of voiced and voiceless onsets and developed a bifurcated vowel system based on the original onset voicing. Old Khmer is considered to be an example of the stage of pre-registrogenesis, while Standard Khmer is placed in the stage of post-registrogenesis.

The Middle Khmer period, considered as a transitional stage between Old Khmer and Standard Khmer, is believed to exhibit a contrastive register distinction emerging as a result of the loss of contrastive voicing in onset position during this period. Specifically, the original voiced onsets in Middle Khmer corresponded to a breathy register, while the original voiceless onsets corresponded to a modal register (Huffman 1985: 141; Ferlus 1992: 61–62; Brunelle & Kirby 2016: 194).

The historical development of Khmer as standardly described is summarized in Table 1.

¹ This work originated as part of the author's dissertation. I am grateful to John Whitman, Sam Tilsen, Abby Cohn, Micheal Weiss, and Pittayawat Pittayaporn for their insightful comments. I also extend my thanks to the audiences at the 32nd Annual Meeting of the Southeast Asian Linguistic Society for their valuable feedback. I would like to thank Paul Sidwell, a Managing Editor at JSEALS, and the anonymous reviewers for their constructive insights. My gratitude also goes to Doug Cooper for sharing the digitized version of Jenner's Pre-Angkorian, Angkorian, and Middle Khmer dictionaries.

Table 1 Historical development of voicing and vowel height contrast in Khmer (Brunelle & Kirby 2016 modified from Huffman 1976)

| | Register | |
|-----------------------|------------|-------------|
| | High | Low |
| Stage 1: voicing | <i>kaa</i> | <i>gaa</i> |
| Stage 2: transitional | <i>káa</i> | <i>k'áa</i> |
| Stage 3: register | <i>káa</i> | <i>kàa</i> |
| Stage 4: vowel height | <i>kaa</i> | <i>kiə</i> |

In Stage 1, which corresponds to Old Khmer (Ferlus 1992: 81), onset voicing was contrastive. Slight vowel perturbation, such as pitch, vowel quality, and voice quality, from onset voicing may have been present. In the second stage, the perturbed features of onset voicing onto the vowel became redundant sub-phonemic features on the following vowels, while the onset may have also undergone partial or complete devoicing. In Stage 3, the language became a register language when the two-stop series merge, and vowel differences, such as pitch, vowel quality, and voice quality, became contrastive. At Stage 4, the language phonologized one phonetic property of register, vowel quality, and lost the other phonetic cues. Khmer then became a language with split vowel system. It is worth noting that the phonetic specifics of each stage in Khmer's historical development remain a topic of debate.

There have been modifications to Huffman's (1976) proposal of Khmer historical stages. Huffman (1976) did not specify when and how the vowel quality changes occurred. Diffloth (1990) proposed an intermediate stage between Huffman's Stage 3 and Stage 4, during which vowel quality differences emerged due to the effect of the vowel's voice quality, which emerged first after the loss of onset voicing contrast. Wayland and Jongman (2002) later refined Diffloth's proposal with a phonetic explanation. They suggested that in Huffman's Stage 2, the onset voicing contrast transitioned into a slack and stiff contrast, leading to the emergence of both voice quality and vowel quality contrasts. Subsequently, the onset slack and stiff contrast was lost after the emergence of both voice quality and vowel quality distinctions. Importantly, both Diffloth (1990) and Wayland and Jongman (2002) noted that their proposed stages may overlap, each stage may encompass a long period, and not all vowels may have undergone the process simultaneously.

In this paper, to gain a comprehensive understanding of registrogenesis in Khmer, I investigate the relative and absolute chronology concerning the loss of onset voicing and the vowel bifurcation, drawing upon philological sources. The chronology of changes was examined through an analysis of rhymes in Khmer poetry (16th - early 19th century) and a graphemic analysis of Khmer inscriptions (late 7th - mid 18th century).

1.1 Research question and hypotheses

This study revolves around the central research question:

- (1) What is the relative chronology of the loss of onset voicing and the vowel bifurcation?

Based on this research question, I have formulated two opposing hypotheses as follows:

- (2) Hypothesis 1: The change in vowel quality preceded the loss of onset voicing in Khmer in terms of chronological order.

If this hypothesis holds true, I would anticipate finding earlier written evidence suggesting the occurrence of vowel quality change compared to evidence-e suggesting the loss of onset voicing in Khmer.

- (3) Hypothesis 2: The loss of onset voicing took place prior to the change in vowel quality, with a distinct gap existing between these two changes. It is assumed that this gap corresponds to a period during which voice quality was contrastive.

If this hypothesis holds true, I expect to find written evidence suggesting the occurrence of vowel quality change appearing later than the evidence indicating the loss of onset voicing in Khmer.

1.2 Background

1.2.1 Dating of historical developments in Khmer language

Historically, stop onsets in proto-Khmer and Old Khmer—spanning both pre-Angkorian and Angkorian periods—were contrastive for voicing (Ferlus 1992; Jacob 1960; 1976). This contrast, however, has been lost in Standard Khmer and its contemporary dialects. Notably, remnants of this voicing contrast can be observed in the split vowel system of Standard Khmer. This paper centers on the transition stages, particularly the loss of onset voicing and the change in vowel quality, as presented between Stages 1 and 4 in Table 1.

Several studies have focused on dating each historical stage of Khmer. A study by Maspong (2022), which analyzed Zhou Daguan’s “Zhenla Fengtu Ji” from the late 13th century, offered insights into Khmer historical phonology. The Chinese-transcribed Khmer words within this account indicate that voiced and voiceless stop onsets were still distinct. Furthermore, the transcriptions hint that vowel bifurcation had yet to materialize by this time. Drawing from this, it is conceivable that Khmer, by the end of the 13th century, might have been positioned in either Stage 1 or Stage 2 of registrogenesis listed in Table 1.

Shedding further light on this chronology, Headley’s (1998) research on Khmer loanwords in Western Cham suggests that during the borrowing phase, Khmer preserved its distinction between voiced and voiceless stops. The evidence is from the fact that the voiced stops in Khmer loanwords merged with native voiced stops and underwent similar changes. What is more, the vowel system of the language had not experienced bifurcation, as the vowel nuclei of these loanwords were preserved with minimal or no alteration. This period of borrowing likely post-dates the mid- to late 15th century schism between Western and Eastern Cham, as the Khmer loanwords are not found in Eastern Cham.

The dating of the devoicing and merger process is commonly hypothesized to be relatively recent according to existing literature. Lewitz’s (1967: 388) investigative work on Khmer toponyms, using 16th to 17th-century Spanish and Portuguese sources (Groslier 1958) as references, posits a shift in voicing between these centuries. Specifically, voiced stops become voiceless ones in the 17th century. Yet, a reanalysis by Vickery in (1992: 244–246), using the same historical materials, contends that the devoicing had already taken place by the 16th century. It is imperative to understand, though, that this evidence only confirms the occurrence of the onset voicing merger, without specifying when the process began.

Finally, pinpointing the exact timeframe for the formation of registers in Khmer is inherently challenging since registers cannot be straightforwardly inferred from written records. Thus, most interpretations lean towards a simultaneous occurrence of register formation with the devoicing and merger of onset voicing events, likely in the 16th century.

The dating of stages in the history of the Khmer language can be summarized as follows:

(4) Summarized historical developments of Khmer

| | | | |
|-----------------------------|-----|-----|---|
| Proto-Khmer (500 AD) | kaa | gaa | (Comparative-historical evidence) |
| Pre-register (1200-1400 AD) | kaa | gaa | (Chinese transcription evidence and Western Cham loanwords) |
| Register (1500 AD) | káa | kàa | (Toponyms from Western sources) |
| Standard Khmer | kaa | kiə | |

1.2.2 Vowel quality change in Khmer

Standard Khmer is known for having a complex vowel system with 31 distinct nuclei, made up of 18 monophthongs and 13 diphthongs (Huffman 1970: 8–11; See also Martini 1946; Henderson 1952; Pinnow 1980 for alternative analyses): 10 long vowels /i: e: ε: i: ə: a: α: u: o: ɔ/, 8 short vowels /i e i ə a α u o/, 10 long diphthongs /iə iə uə ei əi ou ae əə ao əə/, and 3 short diphthongs /ěə ŭə ǒə/. This complex system has its roots in the vowel bifurcation that occurred together with the loss of onset voicing contrast.

Observations indicate that Old Khmer vowels eventually bifurcated, leading to the emergence of the first and second “series” in Standard Khmer. For example, the pre-Angkorian Khmer *a: split into Standard Khmer /a:/ (first series) and /iə/ (second series). Evidence supporting the bifurcation of Khmer vowels is primarily found within the Khmer graphemic system. Khmer features two distinct “series” of consonantal graphemes, each corresponding to a different set of vowel phonemes. Specifically, the pronunciation of the same vowel grapheme varies depending on the series of consonant preceding it. For example, when the first series

consonant grapheme 𑄛 <t> /t/ is followed by the vowel grapheme 𑄛 <ā> (reflecting pre-Angkorian Khmer *a:) resulting in 𑄛 <tā>, it is pronounced as /ta:/. Conversely, when the second series consonant 𑄛 <d> /t/ is followed by the same vowel grapheme, yielding 𑄛 <tā>, it is pronounced as /tiə/, as demonstrated by Huffman (1970).

The consonant inventory of Standard Khmer consists of /p b m w t d s n l r c j k ŋ ʔ h²/ (Huffman 1970: 6). Within this inventory, all voiceless stops correspond to two series of consonant graphemes. For instance, /k/ corresponds to a first series grapheme 𑄛 <k> and a second series grapheme 𑄛 <g>. The complete list of grapheme-phoneme correspondences is detailed in Huffman (1970: 18–20). Historically, first-series consonants were voiceless, with the second-series being voiced. As voicing contrasts vanished in Standard Khmer, as voiced consonants devoiced and merged with voiceless consonants, vowel bifurcation ensued. Following the loss of voicing distinction, Khmer developed a register contrast characterized by vowel quality distinctions. Subsequently, Khmer lost all cues associated with the register complex, retaining only vowel quality distinctions. Ultimately, Khmer evolved into a language with a complex vowel system, featuring additional diphthongs that arose through vowel breaking in specific environments.

A closer examination of vowel quality changes in Khmer reveals the occurrence of vowel mergers for certain vowel sounds. One significant case involves the merger of *a: following original voiced onsets (represented as <ā>) and the diphthong *iə (represented as <ia>) into /iə/. Another similar case is the change of *i: following original voiceless onsets (represented as <ṭ>) to /əj/ and *aj following original voiced onsets (represented as <ai>) into /ij/. Although *i:, which became /əj/, and *aj, which became /ij/, do not exhibit complete merger, some scholars suggested merger in some variants of their pronunciation (Jenner 1973)³, or proposed that the two vowels were pronounced the same at some point in history before diverging to their modern pronunciations (Jacob 1966).

In the subsequent sections, the analyses of Khmer rhymes and Khmer orthography in the inscriptions are presented to extract the chronology of the loss of onset voicing and the vowel bifurcation.

2. Analysis of rhyming in Khmer poetry (16th century to early 19th century)

In this section, the focus is on dating the vowel bifurcation by analyzing rhymes in Khmer poetry from the 16th to the 19th centuries. The study of rhymes in Khmer poetry has been one of the sources in investigating vowel bifurcation in the language (Jenner 1973; 1976; 1977). As Khmer poets use rhymes based on the vowel nuclei and codas of the rhyming syllables, these rhymes reflect the pronunciation of the vowels and final consonants. The underlying assumption is that syllables can only rhyme if their rimes are pronounced the same at the time of composition. If rimes rhyme in poetry and they should have been pronounced differently in Old Khmer but are pronounced the same in the modern language, I can infer that the vowels of the rhyming pair have already merged.

2.1 Predictions

The evolution of vowel quality in Khmer has led to the merging of certain vowels, especially among some monophthongs. When such a merger takes place, one would expect to observe pairs of words that, although they did not rhyme historically, now do in Standard Khmer. For example, the words *ma:n ‘to have’ and *riən ‘to acquire knowledge’ did not rhyme during the Angkorian Khmer era. However, in Standard Khmer, they rhyme as they are now pronounced /miən/ and /riən/, respectively.

Jacob (1966) introduced this phenomenon as “convergent rhymes,” which are orthographically distinct but sound alike in modern pronunciation. It is noteworthy that Khmer’s orthography typically mirrors Old

² It is argued that the consonants /b d/ in Standard Khmer are realized as implosives /b d/, or they exhibit phonological environments where they are realized as implosives. I use the symbols /b d/ here without assuming that the consonants are exclusively voiced stops or implosives.

³ There are various transcriptions of the pronunciations of these graphemes in different sources. The grapheme <ṭ> following first-series consonants has been transcribed as /əj/ by Headley (1977) and Huffman (1970), but as /əəj/ by Jenner (1973), while <aj> following second-series consonants has been transcribed as /ej/ by Headley (1977), as /ij/ by Huffman (1970), and as /əj/ by Jenner (1973). These transcriptions suggest that the two graphemes are different phonemically. However, Jenner (1973) reported that in the actual pronunciations in Standard Khmer of the two graphemes are the same, which is [əj]. In this paper, I follow Huffman’s (1970) transcription.

Khmer pronunciation. This offers a glimpse into the historical phonetic shifts that have occurred. Thus, we can infer that convergent rhymes can be indicative of mergers brought about by changes in vowel quality.

In this paper, we will be concentrating on determining the time frame of these mergers rather than the splits. This emphasis is due to the fact that even in Standard Khmer, the rules against “divergent rhymes” – which are orthographically correct but differ in modern pronunciation (Jacob 1966) – are not rigidly followed in Khmer poetic tradition, even after the completion of vowel bifurcation (Maspong 2023: 54–55). Such divergent rhymes are tolerated in contemporary poetry. Jacob (1966) mentioned that it is uncommon to find poetry devoid of divergent rhymes, with *Diav Ek* (1945) being a notable exception. Hence, when it comes to pinpointing the timeline of the vowel bifurcation process, convergent rhymes are more revealing than divergent ones.

2.2 Methodology

This study delved into ten Khmer literary pieces from the 16th to the 19th century. Four texts from the *Cpā’p* genre—short homiletic writings—were selected based on their clear composition dates, as many in this category are undated (Jenner 1976). Another prominent work, *Rāmakerti*, spanned three centuries and involved multiple authors; its first segment hails from the 16th century and the latter from the 18th century (Jacob 1986). The rest of the chosen texts had a clear date of composition.

For analysis, rhymes were extracted from each piece. Due to the length of some texts, like *Rāmakerti* and *Kākī*, only the initial 500 rhymes were taken to avoid bias. The collection encompassed 3,760 rhymes, roughly a thousand per century, with the 16th century only represented by *Rāmakerti 1*. A detailed list appears in Table 2.

Table 2 List of literature used in this analysis

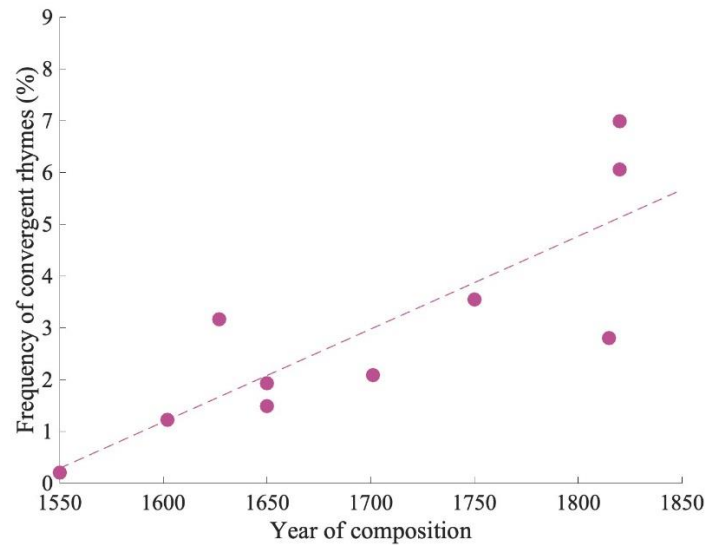
| ID | Text name | Year | # rhyme pairs | Total rhyme pairs |
|----|---|--------------------------------|---------------|-------------------|
| 1 | Rāmakerti 1 | 16 th century | 520 | 520 |
| 2 | Lpoek Angkor Vat | 1602 | 520 | 1,226 |
| 3 | Sarasoer Hemantamās | 1627-1630 | 245 | |
| 4 | Cpā’p Kertikāla | 17 th century | 107 | |
| 5 | Cpā’p Kūn Cau | 17 th century | 354 | |
| 6 | Grande inscription d’Angkor Vat (IMA38) | 1701 | 482 | 1,010 |
| 7 | Rāmakerti 2 | 18 th century | 528 | |
| 8 | Kākī | 1815 | 514 | 1,004 |
| 9 | Cpā’p Prus | early 19 th century | 287 | |
| 10 | Cpā’p Ariyasatthā | early 19 th century | 203 | |
| | | | | 3,760 |

We centered our attention on “convergent rhymes.” As defined by Jacob (1966), these rhymes are distinct in orthography but nonetheless rhyme in contemporary pronunciation. It is posited that such rhymes signify the merging of vowels within the rhyming syllables, as illustrated by the rhyming of Old Khmer *ba: (> Standard Khmer [piə]) and Old Khmer *tiə (> Standard Khmer [tiə]). This perspective follows the analyses presented by Jacob (1966) and Jenner (1976).

2.3 Results

One notable pattern observed is the increasing frequency of convergent rhymes from compositions in the 16th century to the 19th century, which aligns with our expectations considering the documented vowel mergers in later periods of Khmer. Figure 1 illustrates the percentage of convergent rhymes found in each text.

Figure 1 Frequency of convergent rhymes across different texts. The dashed line represents the trendline illustrating the correlation in rhyme frequency over time.



A chronological list of each merger and its associated convergent rhyme pairs is presented below.

2.3.1 Pre-bifurcation merger: *i > i / __ C]_σ

The earliest merger identified is the merger of *i and *i in closed syllables. According to research (Ferlus 1992), this merger is believed to have occurred prior to vowel bifurcation. Confirming this, convergent rhymes reflecting this merger have been found since the 16th century, whereas other convergent rhymes are not yet observed. This merger is evident in the rhyming of <i> and <î>, as demonstrated in (5).

(5) Rhyming of <i> and <î>

| | | | | |
|------|---------------------|--|---|----------------------------|
| C.16 | Rāmakerti 1 | <i>dhammik</i> [-ik > -ik] | : | <i>gagik</i> [-ik] |
| C.17 | Lpoek Angkor Vat | <i>siñh</i> [-iŋ > -iŋ > -əŋ] | : | <i>priñ</i> [-iŋ > -əŋ] |
| | | <i>p'prim</i> [-im > -im > -əm] | : | <i>ñañim</i> [-im] |
| | Sarasoer Hemantamās | <i>niñ</i> [-iŋ > -iŋ] | : | <i>kum.tiñ</i> [-iŋ > -əŋ] |
| | | <i>jit</i> [-it > -it] | : | <i>ñañit</i> [-it] |
| | Cpā'p Kūn Cau | <i>sucarit</i> [-it > -it] | : | <i>briddh</i> [-it] |
| C.19 | Kākī | <i>p'prim</i> [-im > -im > -əm] | : | <i>ñañim</i> [-it] |
| | Cpā'p Prus | <i>git</i> [-it > -it] | : | <i>ñañit</i> [-it] |
| | | <i>brahmagiti</i> [-it > -it] | : | <i>briddh</i> [-it] |

In most cases of rhyming between <i> and <î>, the two syllables rhyme perfectly in modern pronunciation. However, there are instances where they do not rhyme in modern pronunciation, such as the example of *p'prim* : *ñañim*. In this case, the *i of the first rhyme member *p'prim* has lowered to [ə] in Standard Khmer due to vowel bifurcation. It is possible to assume that these rhyme pairs were “perfect” due to a merger that occurred prior to vowel bifurcation and the pairs are “divergent rhymes”.

In summary, it can be safely stated that the merger of *i and *i in closed syllables had already taken place in the 16th century, as convergent rhymes reflecting this merger were found in compositions from that period. This merger occurred prior to the vowel bifurcation. The vowel *i in closed syllables merged with *i before the vowel bifurcation took place, and the two vowels underwent the vowel bifurcation together (Ferlus 1992).

2.3.2 Bifurcation of *u: *u: > [əw] / [-voice] __]σ

Another vowel change that occurred early is the transformation of *u: to [əw] in open syllables with voiceless onsets (Sok 2004)⁴. This change is evident in the rhyming of syllables with <ūv> and <au>, which are pronounced as [aw] following original voiceless onsets or [i̯w] following original voiced onsets in Standard Khmer. The *u: with original voiceless onsets and *aw with original voiced onsets are not fully merged, akin to the situation discussed for *i: and *aj in §2.3.3. Nonetheless, some scholars have suggested merger in their actual pronunciation (Jenner 1973), or hypothesized that the two vowels were pronounced the same at some point in history before diverging to their modern pronunciations (Jacob 1966). Nevertheless, the rhyming is evident the change of *u: to [əw] in open syllables with voiceless onsets.

It is worth noting that in Old Khmer inscriptions, <ūv> was simply written as <ū> to represent *u: in an open syllable, but the addition of <v> possibly reflects a change in pronunciation (Jenner 1974).

(6) Rhyming of <ūv> and <au>

| | | | | |
|------|-------------------|--------------------------|---|--------------------------|
| C.17 | Lpoek Angkor Vat | <i>quv</i> (-u: > -əw) | : | <i>nau</i> (-aw > -i̯w) |
| C.18 | Rāmakerti 2 | <i>truv</i> (-u: > -əw) | : | <i>nau</i> (-aw > -i̯w) |
| | | <i>qīlūv</i> (-u: > -əw) | : | <i>dau</i> (-aw > -i̯w) |
| C.19 | Cpā'p Prus | <i>trūv</i> (-u: > -əw) | : | <i>dau</i> (-aw > -i̯w) |
| | Cpā'p Ariyasatthā | <i>phlūv</i> (-u: > -əw) | : | <i>dau</i> (-aw > -i̯w) |
| | | <i>sñūv</i> (-u: > -əw) | : | <i>dau</i> (-aw > -i̯w) |
| | | <i>qīlūv</i> (-u: > -əw) | : | <i>dau</i> (-aw > -i̯w) |
| | | <i>qīlūv</i> (-u: > -əw) | : | <i>jrau</i> (-aw > -i̯w) |

The rhymes provide evidence that the shift from *u: to [əw] in open syllables with voiceless onsets had already taken place in the 17th century. These rhyme pairs also indicate another change, namely the transformation of *a to [i̯] in syllables with voiced onsets and final glides. This change will be discussed in greater detail in §2.3.4, providing a more comprehensive analysis of this particular vowel shift.

2.3.3 Bifurcation of *i: *i: > [əj] / [-voice] __]σ

Similar to the transformation of *u: to [əw] discussed earlier, *i: in open syllables with voiceless onsets also underwent a change to [əj]. In Standard Khmer, the <ai> grapheme is pronounced as [aj] following original voiceless onsets or [i̯j] following original voiced onsets. It is important to note that <ai> is a grapheme with three allographs: *ai*, *aiy*, and *āy*; and <ᵛ> is a grapheme with three allographs: *ī*, *īy*, and *iy*. Once more, the *i: with original voiceless onsets and *aj with original voiced onsets are not entirely merged. However, these rhyming pairs can still provide evidence of the shift from *i: to [əj] in open syllables with voiceless onsets. This change is reflected in the rhyming of syllables with <ᵛ> and <ai>, as demonstrated in (7).

⁴ Sok (2004) transcribed the result of the change as a short diphthong [əɯ]. However, since they never occur in closed syllables and Khmer has a constraint against complex coda, I transcribe it as a vowel with a coda [əw].

(7) Rhyming of <ɪ̄> and <ai>

| | | | | |
|------|-------------|---------------------------------|---|------------------------------|
| C.18 | IMA38 | <i>śrī</i> (-i: > -əj) | : | <i>didaiy</i> (-aj > -ij) |
| | | <i>snādhivvatiy</i> (-i: > -əj) | : | <i>didaiy</i> (-aj > -ij) |
| | Rāmakerti 2 | <i>qaprīy</i> (-i: > -əj) | : | <i>bhāy</i> (-aj > -ij) |
| | | <i>ksatrī</i> (-i: > -əj) | : | <i>nai</i> (-aj > -ij) |
| | | <i>smī</i> (-i: > -əj) | : | <i>bhāy</i> (-aj > -ij) |
| | | <i>mahesī</i> (-i: > -əj) | : | <i>dīdai</i> (-aj > -ij) |
| | | <i>qaprīy</i> (-i: > -əj) | : | <i>nai</i> (-aj > -ij) |
| | | <i>srī</i> (-i: > -əj) | : | <i>nai</i> (-aj > -ij) |
| | | <i>ksatrī</i> (-i: > -əj) | : | <i>hardāy</i> (-aj > -ij) |
| | | <i>savanīy</i> (-i: > -əj) | : | <i>vināy</i> (-aj > -ij) |
| C.19 | Cpā'p Prus | <i>ktī</i> (-i: > -əj) | : | <i>mai</i> (-aj > -ij) |
| | Cpā'p | <i>ktī</i> (-i: > -əj) | : | <i>suargālāy</i> (-aj > -ij) |
| | Ariyasathā | | | |

The rhymes presented above provide evidence that the shift from *i: to [əj] in open syllables with voiceless onsets had already taken place in the early 18th century. These rhyme pairs also reflect the change of *a to [i] in syllables with a voiced onset and a final glide, which will be discussed in §2.3.4 in greater detail.

2.3.4 Bifurcation of *a: *a > [i] / [+voice] __ [glide]]_σ

Another change that occurred alongside the transformations of *u: to [əw] and *i: to [əj] in open syllables with voiceless onsets is the shift of *a to [i] in syllables with voiced onsets and glide codas. This means that *aw and *aj, represented by the graphemes <au> and <ai>, respectively, became [i^hw] and [ij] following voiced onsets.

There is a debate regarding the original pronunciation of the graphemes <au> and <ai>, questioning whether they were pronounced as *aw and *aj or as *əw and *əj. It is worth noting that *əw and *əj are chosen instead of *i^hw and *ij, because the debate about the original pronunciation is discussed by Jenner (1974), who proposed that the modern pronunciations of the graphemes <au> and <ai> following original voiced onsets are [əw] and [əj], identical to the pronunciation of <ūv> and <ɪ̄> in open syllables following original voiceless onsets.

The proposal that *əw and *əj were the original pronunciations stems from the rhyming patterns observed between <ūv> : <au> and <ɪ̄> : <ai>. Jenner (1974) found that <ūv> and <ɪ̄> following voiceless onsets could rhyme with <au> and <ai> following both voiced and voiceless onsets. Based on this observation, Jenner proposed that the changes of *u: > [əw] and *i: > [əj] occurred when the vowels represented by <au> and <ai> had not yet bifurcated, and thus, the original pronunciations of <au> and <ai> were *əw and *əj. Otherwise, words with <au> and <ai> following original voiceless onsets would not be able to rhyme with *u: and *i: with voiceless onsets. Additionally, Jenner suggested that the bifurcation of *u: and *i: predates the bifurcation of the vowels represented by <au> and <ai>.

(8) Imperfect rhyming of <ūv> : <au> and <ɪ̄> : <ai>

| | | | | |
|------|--------------|---------------------------|---|--------------------|
| C.17 | Lpoek Angkor | <i>qūv</i> (-u: > -əw) | : | <i>phau</i> (-aw) |
| | Vat | | | |
| C.18 | Rāmakerti 2 | <i>qīhūv</i> (-u: > -əw) | : | <i>cau</i> (-aw) |
| | | <i>qīhūv</i> (-u: > -əw) | : | <i>cau</i> (-aw) |
| | | <i>lokiy</i> (-i: > -əj) | : | <i>ksāy</i> (-aj) |
| | | <i>srī</i> (-i: > -əj) | : | <i>visāy</i> (-aj) |
| | | <i>ksatrī</i> (-i: > -əj) | : | <i>phtai</i> (-aj) |
| | | <i>lokiy</i> (-i: > -əj) | : | <i>ksāy</i> (-aj) |

The proposal that <au> and <ai> were pronounced as *əw and *əj in Old Khmer is not without its challenges. One source of counterevidence comes from the rhyming data itself, which includes imperfect rhymes that exhibit divergent rhyming patterns. Such divergent rhymes are not uncommon in Khmer poetry, as discussed in §2.1, where it is noted that divergent rhymes are tolerated even in modern poetry, despite the fact that pronunciations have already diverged (Jacob 1966; Maspong 2023). This suggests that the rhyming of <ūv> and <ᵛ> following voiceless onsets with <au> and <ai> following voiceless consonants, as shown in (8), may not necessarily reflect the relative chronology of vowel changes, but rather the tolerance of Khmer poets for divergent imperfect rhymes.

Additionally, evidence from Khmer loanwords in Thai supports the idea that *aw and *aj were the original pronunciations. Loanwords in Thai with these vowel nuclei are pronounced with [aw] and [aj], rather than [əw] and [əj], although the latter pronunciations are possible rhymes in Thai. This lends support to the notion that *aw and *aj were likely the original pronunciations in Old Khmer.

(9) Khmer loanwords in Thai with <au> and <ai> nuclei

| | Khmer | Thai | |
|---------------|----------|------------------------|--------------|
| <i>khlau</i> | /khlaw/ | /k ^h lǎw/ | ‘foolish’ |
| <i>jamrau</i> | /camrəw/ | /c ^h āmṛāw/ | ‘depth’ |
| <i>cañrai</i> | /cañraj/ | /cāñrāj/ | ‘misfortune’ |
| <i>brai</i> | /prəj/ | /p ^h rāj/ | ‘forest’ |

In conclusion, based on the analysis of rhyming pairs and their patterns, I propose that the original values of <au> and <ai> in Old Khmer were *aw and *aj, respectively. The observed rhyming patterns in (6) and (7) suggest that the change involved the shift of *a to [i] in syllables with voiced onsets and glide codas. These changes likely took place in the 17th century, concurrently with the change of *u: to [əw]. We may also speculate that the shift of *a to [i] preceding a glide coda may have an intermediate stage of *a > [ə] > [i], and the rhyming of <ūv> : <au> and <ᵛ> : <ai> could be considered a “perfect” rhyme in this intermediate stage.

2.3.5 Bifurcation of *e: : *e: > [ae] / [-voice] __

The merger of *e: and *ɛ: into [ae] following voiceless consonants is reflected in the rhyming pairs with <e> and <ae>. It is worth noting that although some words with <e> are written with <ae> in modern spelling, their reconstruction as *e: is supported by the pronunciations in Western Khmer and loanwords in Thai (Ferlus 1992). Indic loanwords like *hetu*, *bises*, and *kiles* with *e: nuclei further support this reconstruction.

(10) Rhyming of <e> and <ae>

| | | | | |
|-----------------------------|------------------------------|-----------------------------|---------------------------------------|--------------------------------|
| C.17 | Lpoek Angkor Vat | <i>qanek</i> (-e:k > -aek) | : | <i>caek</i> (-ɛ:k > -aek) |
| | | <i>hetu</i> (-e:t > -aet) | : | <i>taet</i> (-ɛ:t > -aet) |
| | | <i>qaen̄</i> (-e:ŋ > -aen̄) | : | <i>kansaen̄</i> (-ɛ:ŋ > -aen̄) |
| | | <i>qaen̄</i> (-e:ŋ > -aen̄) | : | <i>plæn̄</i> (-ɛ:ŋ > -aen̄) |
| | | <i>taen̄</i> (-e:ŋ > -aen̄) | : | <i>khvaen̄</i> (-ɛ:ŋ > -aen̄) |
| | | Sarasoer Hemantamās | <i>qaen̄</i> (-e:ŋ > -aen̄) | : |
| <i>qaen̄</i> (-e:ŋ > -aen̄) | : | | <i>ktaen̄</i> (-ɛ:ŋ > -aen̄) | |
| <i>qaen̄</i> (-e:ŋ > -aen̄) | : | | <i>saen̄</i> (-ɛ:ŋ > -aen̄) | |
| Cpā’p Kūn Cau | <i>phqaem</i> (-e:m > -aem) | : | <i>lhaem</i> (-ɛ:m > -aem) | |
| | C.18 IMA38 | <i>kreñ</i> (-e:ŋ > -aen̄) | : | <i>væn̄</i> (-ɛ:ŋ) |
| Rāmakerti 2 | <i>lpaen̄</i> (-e:ŋ > -aen̄) | : | <i>prataen̄</i> (-ɛ:ŋ > -aen̄) | |
| | C.19 Kākī | <i>bises</i> (-e:h > -aeh) | : | <i>kraqaes</i> (-ɛ:h > -aeh) |
| <i>qaen̄</i> (-e:ŋ > -aen̄) | | : | <i>sñaen̄</i> (-ɛ:ŋ > -aen̄) | |
| <i>qaen̄</i> (-e:ŋ > -aen̄) | | : | <i>trataen̄(kaen̄)</i> (-ɛ:ŋ > -aen̄) | |
| Cpā’p Prus | <i>kiles</i> (-e:h > -aeh) | : | <i>prahaes</i> (-ɛ:h > -aeh) | |
| Cpā’p Ariyasatthā | <i>hetu</i> (-e:t > -aet) | : | <i>paek</i> (-ɛ:k > -aek) | |
| | <i>qaen̄</i> (-e:ŋ > -aen̄) | : | <i>svaen̄</i> (-ɛ:ŋ > -aen̄) | |
| | <i>qaen̄</i> (-e:ŋ > -aen̄) | : | <i>taen̄</i> (-ɛ:ŋ > -aen̄) | |

Based on the available data, it is plausible to propose that the change of *e: following voiceless consonants to [ae] occurred in the early 17th century. However, the dating of the change from *ɛ: to [ae] remains uncertain due to the presence of imperfect rhymes, which could be attributed to the tolerance for divergent rhymes, such as *qaen* (-e: > -ae) : *lvaen* (-e:) and *kren* (-e: > -ae) : *vaen* (-ɛ:). It is possible that the *e: following voiceless onsets might have lowered and merged with *ɛ: before *ɛ: bifurcated and then *e: and *ɛ: following voiceless onsets became [ae], i.e., *e: > [ɛ:] > [ae] / [-voice] __ and *ɛ: > [ae] / [-voice] __. In other words, *e: underwent bifurcation before *ɛ:, but further evidence is needed to confirm this claim. Without additional data, I refrain from proposing a specific dating or relative chronology for the bifurcation of *ɛ: based solely on the rhyming data.

2.3.6 Bifurcation of *a: *a > [u] / [+voice] __ [labial]]_o

The analysis conducted in this section reveals the merger of *a and *u in syllables with voiced onsets and final labials [m] and [p] (Ferlus 1992). However, it is important to note that the evidence for this change is found relatively late compared to the other changes discussed in previous sections.

(11) Rhyming of <a> and <u>

| | | | | |
|------|-------------|--------------------------|---|--------------------|
| C.18 | IMA38 | <i>dham</i> (-am > -um) | : | <i>phjum</i> (-um) |
| | | <i>dham</i> (-am > -um) | : | <i>phjum</i> (-um) |
| | Cpā'p Bākya | <i>bhnam</i> (-am > -um) | : | <i>jum</i> (-um) |
| | Cā's | | | |
| C.19 | Kākī | <i>bhnam</i> (-am > -um) | : | <i>jum</i> (-um) |
| | Cpā'p Prus | <i>gra'p</i> (-ap > -up) | : | <i>lup</i> (-up) |
| | | <i>gra'p</i> (-ap > -up) | : | <i>lup</i> (-up) |
| | | <i>dham</i> (-am > -um) | : | <i>num</i> (-um) |

Based on the available data, it is indeed plausible to hypothesize that the merger of *a and *u in syllables with voiced onsets and final labials occurred in the 18th century. Interestingly, the evidence for the merger of the vowel preceding [p] is found later than those preceding [m]. While it is conceivable that these mergers occurred separately, the limited evidence does not provide enough support to confirm or refute this hypothesis. Therefore, as a working hypothesis, it can be assumed that the merger of *a preceding both [m] and [p] took place together around the 18th century.

2.3.7 Bifurcation of *a: *a: > [iə] / [+voice] __

The last merger that I can find evidence of from the rhyming data is the merger of *a: and *iə. This merger involves the change of *a: > [iə] following voiced onsets. Unlike several other diphthongs in Standard Khmer, the diphthong [iə] has been attested since the Old Khmer period or even since proto-Khmer (Ferlus 1992). Unfortunately, there is only one instance of <ā> : <ia> rhyme reflecting this merger, as follow:

(12) Rhyming of <ā> and <ia>

| | | | | |
|------|-------------|---------------------------|---|--------------------|
| C.19 | Cpā'p | <i>brāt</i> (-a:t > -iət) | : | <i>piat</i> (-iət) |
| | Ariyasatthā | | | |

Jenner (1977) proposed that *a:, when following a voiced onset, evolved to [ea] before transitioning to [iə] and subsequently merging with the diphthong [iə] (i.e., *a: > [ea] > [iə]). The late appearance of this rhyme aligns with Jenner's theory for a couple of reasons. First, this rhyme is absent in older texts, despite the relatively common presence of words with *a: and *iə nuclei. Second, in the *Cpā'p Ariyasatthā*, where this rhyming pair emerges, the author seemingly prioritizes pronunciation over the written form. This is further supported by a markedly low presence of divergent imperfect rhymes in this work compared to other texts (Maspong 2023).

In other words, the delayed emergence of such rhymes indicates that the merger between the *a: and *iə nuclei occurred relatively late. Additionally, the bifurcation of *a: must have been established before the onset of the 19th century.

2.4 Summary

The analysis of rhymes in Khmer poetry indicates that vowel bifurcation likely occurred in the 17th century, as reflected by the rhyme pairs that illustrate vowel mergers resulting from this process. Interestingly, the evidence for the bifurcation of each vowel is found in texts from different centuries. This suggests that the mergers, and possibly the bifurcation, did not occur simultaneously for all vowels in the system. The centuries in which evidence for each vowel quality change is found are summarized in Table 3.

Table 3 Chronology of the evidence for vowel change found in rhyming data

| Period | Change | Condition |
|--------------------------|----------------|------------------------------------|
| 16 th century | i > i | __ C] _σ |
| 17 th century | u: > əw | [-voice] __] _σ |
| | a > i | [+voice] __ [glide]] _σ |
| | e: > ae | [-voice] __ |
| 18 th century | i: > əj | [-voice] __] _σ |
| | ɑ > u | [+voice] __ [labial] |
| 19 th century | a: (> ea) > iə | [+voice] __ |

Note that the evidence presented in this study is based solely on the analysis of convergent rhymes. As a result, the findings of this analysis indicate the dating of the (near) merger due to vowel bifurcation, rather than the vowel bifurcation itself.

3. Graphemic analysis of Khmer inscriptions (late 7th century to mid 18th century)

In this section, I explore the dating of the loss of onset voicing distinction and the process of vowel bifurcation by examining spelling variations in Khmer inscriptions. This study covers Khmer inscriptions from the late 7th century, marking the inception of written Khmer records, to the mid-18th century. The underlying premise is that variations in spelling can provide insights into mergers.

3.1 Predictions

In the context of vowel quality changes in Khmer, recall that these changes can lead to both the splitting and merging of vowels. However, for the analysis of the Khmer inscription's graphemic system, we can confidently make predictions mainly about vowel mergers. When a merger occurs within a specific period, we anticipate encountering spelling variations in vowel symbols that were previously used to represent distinct vowel phonemes. To illustrate, consider *ma:n 'to have' and *riən 'to acquire knowledge,' typically written as <mān> and <riən>. If a merger has taken place, resulting in pronunciations like /miən/ and /riən/, we would expect to observe spellings such as <mian> or <rān>.

This same prediction is also employed when considering the merger of stop onsets. It is important to recall that in the Khmer script, the first-series consonant graphemes initially represented original voiceless consonants, while the second-series consonant graphemes represented original voiced consonants. However, following the loss of voicing contrast, both series of consonant graphemes came to represent the same phonemes. For instance, graphemes like <g> and <k>, which once denoted *g and *k respectively, now both represent /k/ in Standard Khmer. To illustrate, take the example of *ki: 'loom' and *gi: 'salt-water shrimp,' typically written as <kī> and <gī> respectively. If a merger occurs, resulting in pronunciations like /kəj/ and /ki:/ respectively, we would anticipate finding spelling variations, such as <gī> for *ki: /kəj/ and <kī> for *gi: /ki:/.

3.2 Methodology

The orthography of words found within Khmer inscriptions have been extracted from three sources: the pre-Angkorian Khmer dictionary (Jenner 2009a), the Angkorian Khmer dictionary (Jenner 2009b), and the Middle Khmer dictionary (Jenner 2011). These dictionaries comprehensively document the spelling variations encountered in Khmer inscriptions, attributing each lexical entry to its respective source and chronological dating to these source materials.

3.3 Results

3.3.1 Evidence for the loss of voicing distinction

The earliest spelling variations of oral stops found is in the late 16th century inscriptions. Such orthographic variation is absent in earlier periods. The examples of these spelling variations are detailed in Table 4.

Table 4 Examples of the first spelling variations of each stop

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|--------------------------|---------------------------------|--------------------|---------------------------|-----------------|
| <i>kusala</i> /kuʔsal/ | <i>kusala</i> ~ <i>gusala</i> | 1578 | <i>kusala</i> /koʔsal/ | ‘merit’ |
| <i>kra</i> /kra:/ | <i>kra</i> ~ <i>gra</i> | 1704 | <i>kra</i> /kra:/ | ‘poor’ |
| <i>guṇa</i> /gun/ | <i>guṇa</i> ~ <i>kuṇa</i> | 1643 | <i>guṇa</i> /kun/ | ‘good deed’ |
| <i>cībara</i> /ci:ba:r/ | <i>cībara</i> ~ <i>jībara</i> | 1671 | <i>cībara</i> /cəjpo:/ | ‘monk’s robe’ |
| <i>jañjīna</i> /jaŋji:ŋ/ | <i>jañjīna</i> ~ <i>cañjīna</i> | 1701 | <i>jañjīna</i> /cūəŋci:ŋ/ | ‘scale’ |
| <i>tvek</i> /tvɛ:k/ | <i>tvek</i> ~ <i>dbaeka</i> | 1704 | <i>thbaek</i> /tpɛ:k/ | ‘to be bald’ |
| <i>dlāya</i> /dla:j/ | <i>dlāya</i> ~ <i>thlāya</i> | 1684 | <i>dhlāya</i> /tliəj/ | ‘to break open’ |
| <i>vleṇa</i> /vlə:ŋ/ | <i>phleṇa</i> ~ <i>bhloṇa</i> | 1578 | <i>bhloṇa</i> /phlə:ŋ/ | ‘fire’ |

The spelling variations indicate that the loss of voicing contrast occurred during the late 16th century. This chronology aligns with dating established through additional evidence, including Khmer loanwords in Western Cham (Headley 1998) and toponyms transcribed by Western missionaries (Vickery 1992).

3.3.2 Evidence for the vowel bifurcation

In addition to the loss of voicing distinction, the spelling variations found within Khmer inscriptions is indicative of vowel mergers. There are a total of six instances of vowel mergers evident in these spelling variations. This dataset further contributes to a more precise chronology of the vowel bifurcation process.

3.3.2.1 Pre-bifurcation: *i > [i] / __ C]σ

The earliest recorded merger, as indicated by both the rhyming data and prior research, pertains to the merger of *i and *ī in closed syllables. It is important to note that the grapheme <ī>, used to denote *ī, is an innovation in Middle Khmer, whereas *ī was formerly represented by <i> in the Old Khmer period. Hence, the appearance of <ī> representing *ī does not inherently imply the merger of *i and *ī. Nevertheless, the usage of <ī> specifically for *ī words implies the merger, as <ī> was specifically created to represent *ī. Although instances of spelling variations among *ī words with both <i> and <ī> are somewhat rare, the earliest variation dates back to an inscription from the late 16th century, as shown in Table 5. Despite the limited data, it is plausible to hypothesize that the merger of *i and *ī had already taken place by the late 16th century.

Table 5 Examples of spelling variation of <i> and <ī>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|-------------------|---------------------------|--------------------|-------------------|------------|
| <i>dika</i> /dik/ | <i>dika</i> ~ <i>dīka</i> | 1599 | <i>dīka</i> /tik/ | ‘water’ |
| <i>gita</i> /git/ | <i>gita</i> ~ <i>gīta</i> | 1697 | <i>git</i> /kit/ | ‘to think’ |

3.3.2.2 Bifurcation of *i: *i (> [i]) > [ə] / [-voice] __

The bifurcation of *i strongly relates to the pre-bifurcation change of *i to [i] discussed in the preceding section. The merger of *i and *i is an intermediate stage in the bifurcation of *i. Following the merger of *i and *i, they underwent the vowel bifurcation process, during which the vowel following a voiceless onset underwent a lowering to [ə]. The evidence of the dating of this lowering is established through the spelling variations in the graphemes <i>, representing *i, and <oe>, representing *ə(:). The earliest instances of these spelling variations are found in the late 17th century, as shown in Table 6, a dating that aligns with expectations, as this change must have occurred subsequent to the merger of *i and *i, for which we have evidence dating back to the 16th century.

Table 6 Examples of spelling variation of <i> and <oe>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|--------------------------|-------------------------------|--------------------|------------------------|---------------------------|
| <i>kārttika</i> /ka:rɨk/ | <i>kātika</i> ~ <i>kātæka</i> | 1684 | <i>kātika</i> /ka:tək/ | ‘the twelfth lunar month’ |
| <i>sthita</i> /sthɨt/ | <i>sthita</i> ~ <i>sthæta</i> | 1684 | <i>sthita</i> /sthət/ | ‘to stay’ |
| <i>tiña</i> /tɨŋ/ | <i>tiña</i> ~ <i>tæña</i> | 1684 | <i>tiña</i> /dəŋ/ | ‘to know’ |

3.3.2.3 Bifurcation of *u: *u: > [o:] / [-voice] __ C]_σ

Another phonological shift reflected in the spelling variations is the lowering of *u: to [o:] in closed syllables with voiceless onsets (Ferlus 1992). The evidence for this change is in the spelling variations between <ū>, representing *u:, and <o>, representing *o:. Direct evidence in the form of spelling variation of words originally containing the vowel *u: is absent. However, indirect evidence is available: certain words originally having the vowel *o: are written with the grapheme <ū>. The underlying presumption is that if the vowel underwent bifurcation, the grapheme <ū> would represent two distinct phonemes, namely [u:] and [o:]. More precisely, <ū> would denote [o:] if it followed original voiceless consonants. Thus, the presence of some instances of words with *o:, specifically those with original voiceless onsets, transcribed as <ū> would indicate that the merger had already occurred.

The earliest instances of *o: written as <ū> date back to the late 16th century. Consequently, it can be inferred that the shift from *u: to [o:] in syllables with voiceless onsets had already occurred during the late 16th century.

Table 7 Examples of spelling variation of <ū> and <o>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|---------------------------|---------------------------------|--------------------|----------------------------|-----------------|
| <i>qoya</i> [ʔo:y] | <i>qoya</i> ~ <i>qūya</i> | 1579 | <i>qoya</i> [ʔaoy] | ‘to give’ |
| <i>qoñkāra</i> [ʔo:ŋka:r] | <i>qoñkāra</i> ~ <i>qūñkāra</i> | 1599 | <i>qoñkāra</i> [ʔao:ŋka:r] | ‘royal command’ |

3.3.2.4 Bifurcation of *i: and *a: *i: > [əj] / [-voice] __]_σ and *a > [i] / [+voice] __ j]_σ

From the rhyming data, it has become apparent that the change involving *i: in open syllables with voiceless onsets and *a in syllables with voiced onsets and a *j coda are closely interconnected. The expected evidence for this relationship in the spelling is the spelling variations between <ɨ>, representing *i:, and <ai>, representing *aj. However, such instances are rare, primarily due to the distinction that <ɨ> signifies the phoneme [əj] in Standard Khmer only when it follows an original voiceless consonant, while <ai> denotes [ij] only when following an original voiced consonant.

The evidence found is indirect evidence. The single instance of spelling variation between <ɨ> and <ai> in words featuring the original vowel *aj dates back to the late 17th century. This aligns with the findings from the rhyming data, supporting the notion that these changes had already occurred by the 17th century.

Table 8 Spelling variation of <ɨ> and <ai>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|------------------|---------------------------|--------------------|-------------------|-----------|
| <i>jai</i> /jaj/ | <i>jāya</i> ~ <i>jīya</i> | 1684 | <i>jāya</i> /cɨj/ | ‘victory’ |

3.3.2.5 Bifurcation of *u: and *a : *u: > [əw] / [-voice] __]_σ and *a > [i] / [+voice] __ w]_σ

The last set of changes for which I have found evidence in the inscription data pertains to the change of *u: in open syllables with voiceless onsets and the change of *a in syllables with voiced onsets and *w codas. The anticipated spelling variation involves <ū> or <ūv>, representing *u:, and <au>, representing *aw. As in the previous section, such variations are scarce due to the non-overlapping distributions of <ūv> and <au>, representing the resultant vowel [əw] and [iɰ] respectively.

The sole evidence of this merger emerges from the spelling variation of a word with the original vowel *aw. This evidence, too, is traced back to the late 17th century, coinciding with the merger of *i: and *aj discussed in the preceding section.

Table 9 Spelling variation of <ū(v)> and <au>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|-----------|--------------|--------------------|--------------|-------|
| nau [naw] | nau ~ nūva | 1684 | nau [niɰ] | ‘new’ |

Additional evidence supporting the shift of *u: in open syllables with voiceless onsets comes from Jenner (1977). He suggested that the alteration in the spelling of [u:] in open syllables from <ū> to <ūv> corresponds to the change in pronunciation to [əw], with the final <v> signifying the presence of the [w] sound. Consequently, I conducted a survey to ascertain the earliest instances of variation between <ū> and <ūv>. My findings reveal that this spelling variation begins to surface in the first half of the 17th century, aligning with the conclusions drawn from the rhyming data.

Table 10 Examples of spelling variation of <ū> and <ūv>

| Old Khmer | Middle Khmer | First found (year) | Modern Khmer | Gloss |
|----------------|-----------------|--------------------|------------------|------------------|
| trū /tru:/ | trū ~ trūva | 1631 | trūva /trəw/ | ‘correct’ |
| phlū /phlu:/ | phlū ~ phlūva | 1684 | phlūva /phləw/ | ‘street’ |
| thkū /thku:/ | thkū ~ thkūva | 1750 | thkūva /thkəw/ | ‘a kind of tree’ |
| satrū /satru:/ | satrū ~ satrūva | 1704 | satrūva /satrəw/ | ‘enemy’ |

4 Discussion and conclusion

4.1 Discussion

In this study, I investigated both the relative and absolute chronologies concerning the loss of onset voicing contrast and vowel bifurcation, drawing upon philological evidence.

As a result, I did not uncover evidence suggesting the loss of voicing distinction or vowel bifurcation in Khmer documents predating the 16th century. Notably, the first spelling variations for initial stops emerged in a Khmer inscription from 1578, and similar variations were evident in subsequent inscriptions. This suggests that the distinction between voiced and voiceless stops had either already merged by this time, or, at the very least, that the process of onset voicing contrast loss was underway between the early to mid-16th century.

The earliest evidence hinting at vowel bifurcation, specifically the bifurcation of *u: in closed syllables, emerged in the late 16th century. This evidence takes the form of spelling variations found in an inscription dated to 1579. Furthermore, I observed that evidence for the bifurcation of high vowels precedes that of other vowels.

The timeline detailing all vowel changes is presented in Table 11.

Table 11 Chronology of the evidence for vowel change

| Period | Change | Condition | Evidence |
|---------------------------------|----------------|------------------------------------|-----------------------------|
| (late) 16 th century | i > i | __ C] _σ | Rhyming, Inscription |
| | u: > o: | [-voice] __ C] _σ | Inscription |
| 17 th century | u: > əw | [-voice] __] _σ | Rhyming, Inscription |
| | i > ə | [-voice] __ | Inscription |
| | i: > əj | [-voice] __] _σ | Rhyming (C.18), Inscription |
| | e: > ae | [-voice] __ | Rhyming |
| | a > ə | [+voice] __ [glide]] _σ | Rhyming, Inscription |
| 18 th century | ɑ > u | [+voice] __ [labial] | Rhyming |
| 19 th century | a: (> ea) > iə | [+voice] __ | Rhyming |

Interestingly, evidence for changes in high vowels surfaced earlier than for low vowels. If, as the evidence suggests, the vowel bifurcation affected high vowels before low vowels, this pattern could offer deeper insights into the process of vowel bifurcation. It is also observed in Chanthaburi Khmer that vowel bifurcation occurred to all short vowels except a low vowel *a, which is realized as [a] for high register and [a̠] (Wayland 2001). However, note that vowel /a/ between low and high register is phonetically different in vowel height indicated by the lower F1 (higher vowel) for low register [a̠] than the F1 of high register [a]. Furthermore, the difference in F1 is comparable with the difference between *i which is realized as [i̠] in low register and [e] in high register.

Interestingly, evidence for changes in high vowels surfaced earlier than for low vowels. If, as the evidence suggests, the vowel bifurcation affected high vowels before low vowels, this pattern could offer deeper insights into the process of vowel bifurcation. This pattern is also observed in Chanthaburi Khmer, as vowel bifurcation occurred for all short vowels except for the low vowel *a, which is realized as [a] in the high register and [a̠] in the low register (Wayland 2001). However, it is important to note that, even though categorical bifurcation is not observed for the low vowel, the vowel /a/ exhibits phonetic differences in vowel height, as indicated by the lower F1 (indicating a higher vowel) for the low register [a̠] compared to the F1 of the high register [a]. Furthermore, the difference in F1 is comparable to the difference between *i, which is realized as [i̠] in the low register, and [e] in the high register (Wayland & Jongman 2001).

Nonetheless, the different chronology of changes for vowels with different qualities aligns with the proposals made by Diffloth (1990) and Wayland and Jongman (2002), suggesting that vowels may not have undergone bifurcation simultaneously. However, there is a limitation to the analyses. Both the rhyme analysis in Khmer poetry and the spelling variations in Khmer inscriptions primarily highlight the presence of mergers (or near mergers) in a given period. These mergers only relate to a subset of vowels in the bifurcation process. Thus, while my analysis can pinpoint when certain vowel qualities, altered due to registrogenesis, merged with other existing vowels, it cannot definitively state when each vowel bifurcation occurred.

By amalgamating evidence from Khmer poetry and Khmer inscriptions, I found evidence of both the loss of voicing distinction and the vowel bifurcation within documents from roughly the same period. However, the sequence in which these phenomena occurred remains ambiguous.

These observations are consistent with prior research. For instance, an examination of a Chinese travel account (Maspong 2022) and an investigation into Khmer loanwords in Western Cham (Headley 1998) both indicated that, during the 13th and 15th centuries respectively, Khmer still retained a contrastive onset voicing and had not yet undergone vowel bifurcation. In contrast, a study analyzing Khmer word transcriptions by Spanish and Portuguese missionaries proposed that the loss of onset voicing contrast had already occurred by the 16th century (Vickery 1992).

The analysis of Khmer philological data in this paper and the previous literature suggest that both the loss of onset voicing and the vowel bifurcation had taken place by the late 16th century. However, I cannot conclusively determine whether the loss of onset voicing preceded the vowel bifurcation or if the reverse occurred. Based on the results, I can infer that the interval between the two events was likely minimal, perhaps less than a century, if there was any discernible gap at all.

4.2 Conclusion

In this study, I endeavored to establish both an absolute and a relative chronology for the loss of onset voicing and vowel bifurcation, key components of Khmer's registrogenesis. My analysis hinged on two primary philological sources: (i) Khmer poetry spanning the 16th to the 19th centuries, and (ii) Khmer inscriptions ranging from the late 7th to the mid-18th centuries. The findings indicate that both changes were in motion by the late 16th century. However, the precise sequential order of these shifts remains ambiguous.

References

- Brunelle, Marc & James Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistics Compass* 10.4:91–207. <https://doi.org/10.1111/lnc3.12182>
- Diffloth, Gérard. 1990. A history of the Khmer language. Unpublished Work. Ithaca, NY, ms.
- Ferlus, Michel. 1992. Essai de phonétique historique du khmer (du milieu du premier millénaire de notre ère à l'époque actuelle). *Mon-Khmer Studies* 21:57–89.
- Groslier, Bernard-Philippe. 1958. *Angkor et le Cambodge au XVIe siècle d'après les sources portugaises et espagnoles* (Annales Du Musée Guimet). Paris: Presses Universitaires de France.
- Headley, Robert K. 1977. *Cambodian-English dictionary*. Washington: Catholic University of American Press.
- Headley, Robert K. 1998. *Cham evidence for Khmer sound changes* (Papers in Southeast Asian Linguistics No.15: Furthur Chamic Studies). Canberra: Pacific Linguistics.
- Henderson, Eugénie J. A. 1952. The Main Features of Cambodian Pronunciation. *Bulletin of the School of Oriental and African Studies* 14.1:149–174. <https://doi.org/10.1017/S0041977X00084251>
- Huffman, Franklin E. 1970. *Cambodian system of writing and beginning reader*. Yale University Press.
- Huffman, Franklin E. 1976. The register problem in fifteen Mon-Khmer languages. In Philip N Jenner & et.al (eds.), *Austroasiatic Studies I (Oceanic Linguistics Special Publication 13)*, 575–589. Honolulu: University of Hawaii Press.
- Huffman, Franklin E. 1985. Vowel permutation in Austroasiatic languages. In Graham Thurgood, James A. Matisoff & David Bradley (eds.), *Linguistics of the Sino-Tibetan area: The state of the art*, 141–144. Canberra: Australian National University Printing Service.
- Jacob, J. M. 1960. The Structure of the Word in Old Khmer. *Bulletin of the School of Oriental and African Studies, University of London* 23.2:351–368.
- Jacob, Judith M. 1966. Some features of Khmer versification. In *In memory of J.R. Firth*, 227–242. London: Longmans.
- Jacob, Judith M. 1976. An examination of the vowels and final consonants in correspondences between Pre-Angkor and Modern Khmer. In Nguyen Dang Liem (ed.), *Southeast Asian Linguistics Studies*, vol. 2, 19–38. Canberra: Australian National University Press.
- Jacob, Judith M. (ed.). 1986. *Reamker (Ramakerti) the Cambodian version of the Ramayana*. London: Royal Asiatic Society.
- Jenner, Philip N. 1973. The value of au and ai in Middle Khmer. In Nguyen Dang Liem (ed.), *Southeast Asian Linguistic Studies* (Pacific Linguistics), vol. 31, 157–173. Canberra: Australian National University.
- Jenner, Philip N. 1974. The development of the registers in Standard Khmer. In Nguyen Dang Liem (ed.), *Southeast Asian Linguistics Studies*, vol. 1, 47–60. Canberra: Australian National University Press.
- Jenner, Philip N. 1976. The relative dating of some Khmer Cpā'pa. In *Oceanic Linguistics Special Publications, No. 13, Austroasiatic Studies Part II*, 693–710. Honolulu: University of Hawai'i Press.
- Jenner, Philip N. 1977. Jenner: The value of i, ī, u, and ū in Middle Khmer. *Mon-Khmer Studies* 5:101–133.
- Jenner, Philip N. 2009a. *A dictionary of pre-Angkorian Khmer*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Jenner, Philip N. 2009b. *A dictionary of Angkorian Khmer*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

- Jenner, Philip N. 2011. *A dictionary of Middle Khmer*. Canberra: Pacific Linguistics.
- Lewitz, Saveros. 1967. La toponymie khmère. *Bulletin de l'École française d'Extrême-Orient* 53.2:375, 377–451.
- Martini, François. 1946. Aperçu phonologique du cambodgien. *Bulletin de la Société de linguistique de Paris* 42:113–131.
- Maspong, Sireemas. 2022. Khmer onset voicing at the end of the 13th century: the Zhēnlà Fēngtǔ Jì revisited. *Cahiers de Linguistique Asie Orientale* 51:175–208.
- Maspong, Sireemas. 2023. Synchrony and diachrony of vowel quality difference across registers in Mon-Khmer languages. Cornell University Ph.D. dissertation.
- Pinnow, Heinz-Jürgen. 1980. Reflection on the history of the Khmer phonemic system. *Mon-Khmer Studies* 8:103–130.
- Sok, Phal. 2004. *Développement des diphtongues en khmer moderne*. Université de Montréal Thesis.
- Vickery, Michael. 1992. Loan words and devoicing in Khmer. *Mon-Khmer Studies* 18–19:240–250.
- Wayland, Ratee P. 2001. Acoustic and perceptual investigation of breathy and clear phonation in Chanthaburi Khmer: Implications for the history of Khmer phonology. Cornell University Ph.D. dissertation.
- Wayland, Ratee P. & Allard Jongman. 2001. Chanthaburi Khmer vowels: phonetic and phonemic analyses. *Mon-Khmer Studies* 31:65–82.
- Wayland, Ratee P. & Allard Jongman. 2002. Registrogenesis in Khmer: A phonetic account. *Mon-Khmer Studies* 32:101–115.

Reviewed: Received 19 January 2024, revised text accepted 14 March 2024, published 1 May 2024
Editors: Editor-In-Chief Dr Mark Alves | Managing Eds. Dr Paul Sidwell, Dr Nathan Hill, Dr Sigrid Lew