

สรุปประเด็นข้อเสนออักษรอีสาน

เทพพิทักษ์ การุญบุญญานันท์

มกราคม ๒๕๕๗

๑ สรุปประเด็นอักษรธรรมลาว-อีสาน

๑.๑ อักขระที่ต้องการเพิ่ม

ตามข้อกำหนดยูนิโค้ด อักษรธรรมลาว-อีสานสามารถใช้รหัสร่วมกับอักษรธรรมอื่นได้แก่อักษรธรรมล้านนาและจีน แต่เมื่อได้สำรวจเอกสารโบราณแล้ว พบว่ายังขาดอักขระอักษรธรรมลาว-อีสานที่ยังไม่มีในตารางอีกสองตัว ได้แก่ ก สะกดจุดคู่ และ สระแอลดรูป

๑.๑.๑ ก สะกดจุดคู่

รูปร่าง: จุดคู่แนวนอนวางใต้พยัญชนะต้น

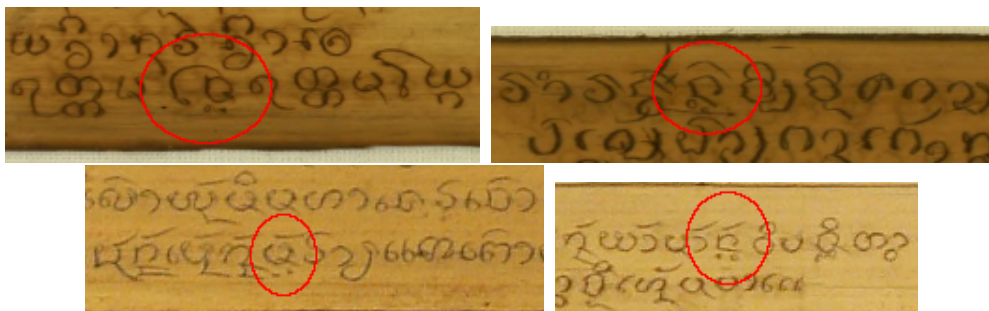
อักขระคล้าย: U+0324 COMBINING DIAERESIS BELOW

หน้าที่: เสียงสะกดแม่กในพยางค์

ในบทเรียนอักษรธรรมที่มีในภาคอีสานของไทย มี ๔ วิธีที่จะเขียนตัวสะกดแม่กในพยางค์ ได้แก่:

๑. ใช้ ก ตัวเต็ม (๓) เมื่อไม่มีที่ว่างทั้งใต้และเหนือพยัญชนะต้น เช่น ๓๓ (หลัก), ๓๓ (หยิก), ๓๓ (หมึก)
๒. ใช้ ก เพียง (๓) เมื่อมีที่ว่างใต้พยัญชนะต้น เช่น ๓๓ (จิก), ๓๓ (ปึก), ๓๓ (หลอก)
๓. ใช้ไม้ขีด (๓) เมื่อมีที่ว่างเหนือพยัญชนะต้น เช่น ๓๓ (มาก), ๓๓ (ออก), ๓๓ (หมวก), ๓๓ (ทุก)

๔. ใช้จุดคู่ใต้พยัญชนะต้น (..) เมื่อมีที่ว่างใต้พยัญชนะต้น คล้ายกรณีใช้ ก เพียง เช่น โฉ (โชค), ฉ (นก), ฉ (มัก), ฉ (นั้ก)



รูปที่ ๑: ตัวอย่าง ก สะกดจุดคู่จากใบลาน

อย่างไรก็ดี ในเอกสารที่สำรวจ ไม่พบการใช้ ก สะกดจุดคู่ในบทเรียนหรือเอกสารโบราณจากฝั่ง สปป. ลาวเลย ทางออกที่เป็นไปได้:

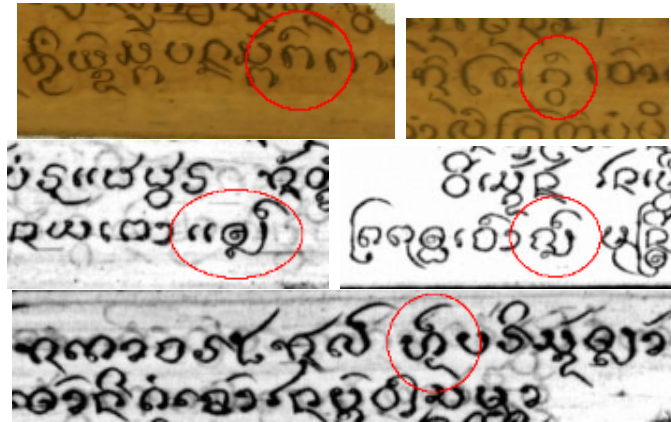
- ใช้อักขระ U+0324 COMBINING DIAERESIS BELOW อย่างน้อยก็เป็นการชั่วคราว จนกว่าจะมีอักขระในช่วงอักษรธรรมโดยเฉพาะให้ใช้ หรืออาจจะเป็นทางออกถาวรก็ได้ แล้วแต่ความเหมาะสมทางเทคนิค
- เพิ่มอักขระยูนิโค้ดในช่วงอักษรธรรม โดยช่องที่ยังว่างและน่าจะเป็นไปได้คือ U+1A5F ซึ่งอยู่ท้ายหมวดอักขระตัวเพื่องพิเศษพอดี

๑.๑.๒ สระแอลดรูป

รูปร่าง: หัวหมวดตัวัดหางเป็นวงโค้งนูนขึ้นด้านบน คล้ายไม้หันอากาศว่า
 อักขระคล้าย: U+1A6B TAI THAM VOWEL SIGN O
 หน้าที่: สระแอ

คำว่า “แอ” สามารถเขียนได้ในรูปของ ล ลิง กับสระแอลดรูป ซึ่งบางที่รูปนี้จะเชื่อมรวมกันกลายเป็นอักขระ ฉิ (U+1A53 TAI THAM LETTER LAE)

แต่สระแอลดรูปนี้ยังมีใช้ในคำอื่นๆ อีก ทั้งในบทเรียนและเอกสารโบราณ ทั้งจากฝั่งไทยและลาว เช่น ฉิ (แก), ฉิ (แก้ว), ฉิ (แล้ว), ฉฉิ (แก้แล), ฉฉิ (ชะแล), ฉฉิ (แห่ง)



รูปที่ ๒: ตัวอย่างสระแอลดรูปจากโบราณ

สระแอลดรูปนี้ มีรูปร่างคล้ายไม้แก้ง (◌) แต่เมื่อสังเกตบทเรียนและเอกสารโบราณจำนวนหนึ่ง จะพบว่าหลายฉบับแยกความแตกต่างอย่างชัดเจน โดยไม้แก้งจะไม่มีหัวขมวด แต่สระแอลดรูปจะมีหัวขมวด

ทางออกที่เป็นไปได้:

- เพิ่มอักขระยูนิโค้ดในช่องอักขรธรรม โดยช่องที่ยังว่างและน่าจะเป็นไปได้คือ U+1A7D หรือ U+1A7E แต่ตำแหน่งที่เหมาะสมที่สุดน่าจะอยู่ต่อจากสระแอลปกติ (U+1A6F) แต่อักขระช่วงนั้นได้ถูกใช้เต็มหมดแล้วตั้งแต่ U+1A70 จนถึง U+1A7C

ตัวร่างเอกสาร: <http://linux.thai.net/~thep/esaan-scripts/tham-issues/new-chars.html>

๑.๒ ไม้อังแล่น/ไม้แก้งไหล

ไม้อังแล่นหรือไม้แก้งไหล (◌) เป็นเครื่องหมายแทนเสียง ง สะกดในภาษาบาลีที่มีพยัญชนะวรรคกะซ้อนตาม โดยคำว่า “แล่น” หรือ “ไหล” ในชื่อ สื่อถึงการเลื่อนตำแหน่งไปวางไว้บนพยัญชนะตัวถัดไปที่ซ้อนตามมา เช่น ฅฎิระ (มงคลละ), มๅฅวิๅ (สงโฆ), มๅฅวิๅ (สงกราส = ศักกราช)

อย่างไรก็ดี ในระบบการเขียนสมัยใหม่ของอักษรธรรมอื่น การ “ไหล” ดังกล่าวเริ่มจะแปรเปลี่ยนไป เนื่องจากตามหลักตรรกะของการสะกดแล้ว เสียง ง มาก่อนเสียงพยัญชนะตาม จึงสะดวกที่จะเขียนก่อนพยัญชนะตามมากกว่าจะเขียนพยัญชนะ

ตามแล้วค่อยเขียนไม้ก่งไหลซ้อนข้างบน ทำให้ตำแหน่งของไม้ก่งไหลเปลี่ยนมาอยู่ที่พยัญชนะต้นแทน ปรากฏการณ์นี้เกิดขึ้นกับอักษรธรรมอื่นที่ไม่ใช่อักษรธรรมลาว-อีสาน ได้แก่อักษรธรรมล้านนาและจีน ทำให้เกิดรูปแบบการเขียนที่แตกต่างกัน และเนื่องจากอักษรธรรมทั้งหมดใช้รหัสยูนีโค้ดร่วมกัน การกำหนดวิธีลงรหัสและแสดงผลจึงต้องพยายามรองรับทุกรูปแบบดังกล่าวในเวลาเดียวกัน

พฤติกรรมการไหล พอจะสรุปได้เป็น ๓ รูปแบบ ดังนี้:

๑. ไหลเสมอ โดยจะเขียนไม้ก่งไหลไว้บนพยัญชนะตามเสมอโดยไม่มีข้อยกเว้น ซึ่งน่าจะเป็นวิธีเขียนตามแบบโบราณ เช่น ພູ້ະ (มุงคะ), ມະໝຸ (สงโหม), ມະໂງູ (สงกราส = ศักราช), ມະໝີ (สงหม)

การไหลแบบนี้พบได้ในอักษรธรรมลาว-อีสานทั้งหมด และในอักษรธรรมล้านนาแบบเก่า

๒. ไม้ไหล หรือเหมือนกับไม้ไหล โดยจะเขียนไม้ก่งไหลไว้บนพยัญชนะต้น โดยอาจจะเขียนเยื้องไปทางขวาบ้าง และเมื่อมีอักขระมาคั่นกลาง (เช่น ในคำว่า ມະໝຸ (สงโหม) ซึ่งมีไม้เกมาคั่นกลาง) ก็จะไม่เลื่อนไม้ก่งไหลข้ามอักขระนั้น โดยตรรกะจึงเหมือนกับเขียนกำกับพยัญชนะต้น ไม่ใช่พยัญชนะตาม

การไหลแบบนี้พบได้ในอักษรธรรมจีน และในหนังสือล้านนาบางส่วน

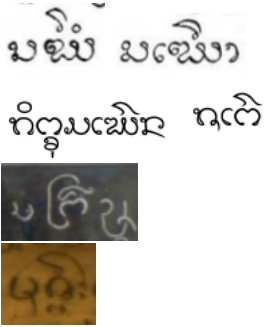
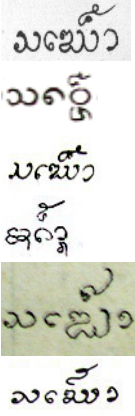
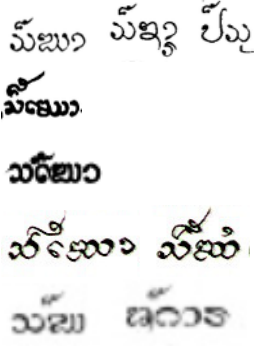
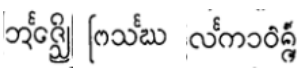
๓. ไหลแบบมีเงื่อนไข โดยปกติจะไหลไปบนพยัญชนะตาม (เช่น ວູ້ວູ້ (องคาร)) แต่จะไม่ไหลเมื่อมีอักขระมาคั่นกลาง (เช่น ມີໂງູ (สงกรานุต) ซึ่งมีระวางมาคั่น) หรือไม่มีที่ว่างบนพยัญชนะตาม (เช่น ຮີວີ (รงสี) ซึ่งที่ว่างบน ສ ถูกไม้กักจับจองแล้ว)

การไหลแบบนี้พบได้ในหนังสือล้านนาบางฉบับ เช่น พจนานุกรมของมหาวิทยาลัยแม่ฟ้าหลวง

ดูสรุปตัวอย่างการเขียนได้ในรูปที่ ๑.๒

กล่าวโดยสรุป อักษรธรรมลาว-อีสานมีความแน่นอนว่าไม้ก่งไหลจะไหลเสมอ อักษรธรรมจีนก็มีความแน่นอนว่าไม้ก่งไหลจะไม่เคยไหล แต่อักษรธรรมล้านนาจะมีความแปรปรวนของรูปแบบมากที่สุด โดยสามารถเป็นไปได้อย่างทั้งสามรูปแบบ ขึ้นอยู่กับสำนักทางออกที่เป็นไปได้:

๑. ใช้รหัสต่างกัน โดยในรูปแบบที่ไหล ก็ให้แทรกพินทุหรือ SAKOT (U+1A60) ระหว่างไม้ก่งไหลกับพยัญชนะตาม เช่น คำว่า สงโหม ก็ให้ลงรหัสเป็น <HIGH

พฤติกรรม	ลาว-อีสาน	ล้านนา	จีน
ไหลเสมอ			-
ไม่ไหล	-		
มีเงื่อนใจ	-	(ไม่มีภาพตัวอย่าง)	-

รูปที่ ๓: เปรียบเทียบการใช้ไม้ก้ำไหลในอักษรธรรมลาว-อีสาน, ล้านนา, จีน

SA, MAI KANG LAI, SAKOT, LOW KHA, VOWEL E, VOWEL AA> ส่วนรูปที่ไม่ไหล ก็เพียงแต่ไม่แทรก SAKOT เท่านั้น

เหตุผล: เป็นรูปแบบการสะกดที่ต่างกัน และอาจปรากฏในเอกสารเดียวกันได้

วิธีการ: rendering engine ตรวจจับลำดับดังกล่าว และสลับลำดับให้

ปัญหา: ตามข้อกำหนดในเอกสาร ISO/IEC JTC1/SC2/WG2 N3207R การลงรหัสของคำว่า “ทั้งหลาย” ในรูปที่ใช้ไม้ก้ำไหล คือ ฎัฎฐ จะมีลำดับเป็น <LOW TA, MAI KANG LAI, SAKOT, LA, VOWEL AA, SAKOT, LOW YA> ซึ่งจะมีลำดับ <MAI KANG LAI, SAKOT, LA> ที่หากเลือกวิธีการนี้จะทำให้ rendering engine ตีความผิดว่าต้องการไม้ก้ำไหลวางบน ล

ตัวเต็ม ทั้งที่ความจริงต้องการให้ <SAKOT, LA> รวมกันเป็น ล เพียงข้าง

๒. ใช้รหัสเดียวกัน โดยทั้งรูปแบบที่ไหลและไม่ไหลควรจะลงรหัสเหมือนกัน และให้ฟอนต์เป็นตัวกำหนดว่าจะวาดแสดงแบบไหน

เหตุผล: การวางตำแหน่งที่ต่างกันของไม้ก้ำไหลไม่ได้ทำให้ความหมายของคำเปลี่ยนไป แต่เป็นเรื่องของความนิยมเสียมากกว่า

วิธีการ: มีความเป็นไปได้คือ

- ฟอนต์ที่ต้องการให้ไหล (กล่าวคือฟอนต์อักษรธรรมลาว-อีสาน และฟอนต์อักษรธรรมล้านนาตามแบบโบราณ) อาจสร้างกฎ GSUB เพื่อสลับลำดับไม้ก้ำไหลกับพยัญชนะตาม ซึ่งจำเป็นต้องให้ rendering engine อนุญาตให้ใช้กฎ GSUB ข้าม cluster ได้ และอาจทำให้เกิดการวางเคอร์เซอร์ที่ไม่เป็นธรรมชาติได้
- ฟอนต์ที่ต้องการให้ไหล อาจใช้ฟิเจอร์ ‘rphf’ ของ OpenType ในลักษณะเดียวกับ Kinzi ของพม่า เพื่อส่งสัญญาณให้ rendering engine สลับลำดับไม้ก้ำไหล

ปัญหา: ความแตกต่างระหว่างรูปแบบ ๑ (ไม่ไหล) กับรูปแบบ ๓ (ไหลแบบมีเงื่อนไข) จะซับซ้อนเกินกว่าจะแทนด้วยวิธีการดังกล่าวได้ และถึงแม้จะใช้ฟิเจอร์ ‘rphf’ rendering engine ก็ยังไม่มีข้อมูลเพียงพอสำหรับแยกแยะระหว่างรูปแบบ ๑ และ ๓ ได้อยู่ดี

ประเด็นเรื่องไม้อังแล่น/ไม้ก้ำไหลนี้ ขณะนี้ยังไม่มีข้อสรุป ยังอยู่ระหว่างสำรวจความเป็นไปได้ทางเทคนิค

ตัวอย่างเอกสาร: <http://linux.thai.net/~thep/esaan-scripts/tham-issues/mai-kang-lai.html>

๒ สรุปประเด็นอักษรไทน้อย

๒.๑ อักขระที่ต้องการเพิ่ม

อักษรไทน้อยคืออักษรที่ฝั่ง สปป. ลาว ได้วิวัฒนาการมาเป็นอักษรลาวปัจจุบัน โดยอักษรไทน้อยครอบคลุมอักษรลาวปัจจุบันทุกตัว และมีอักษรโบราณบางชุดเพิ่มเติมที่ปัจจุบันเลิกใช้ไปแล้ว ในเมื่อปัจจุบันอักษรลาวได้รหัสยูนิโคดในช่วง U+0E80 ถึง

U+OEFF อยู่แล้ว จึงสมเหตุสมผลที่จะขอเพิ่มอักขระต่อเติมจากอักษรลาวปัจจุบัน เพื่อใช้แทนอักษรไทน้อย

โดยหลักการแล้ว ตารางยูนิโค้ดของอักษรลาวถอดแบบมาจากอักษรไทย โดยเว้นช่องอักขระที่มีในอักษรไทยแต่ไม่มีในอักษรลาวไว้ และในอดีต อักษรลาวก็เคยมีอักษรในช่องเหล่านี้ใช้ ดังนั้น การขอเพิ่มอักขระจึงไม่ควรจะต้องช่องว่างเหล่านี้ เพราะอาจมีความต้องการใช้อักขระเหล่านี้ในอนาคตได้ นอกจากนี้ การเว้นไว้ก็ยังทำให้การหีบยืม implementation ระหว่างอักษรไทยและลาวเป็นไปได้โดยสะดวกอีกด้วย

จากการสำรวจบทเรียนอักษรไทน้อยและเอกสารโบราณ พอได้ข้อสรุปของอักษรที่ต้องการเพิ่มดังนี้:

- สระออย (ງ) อักษรลาวปัจจุบันไม่ใช่แล้ว แต่อักษรไทน้อยจำเป็นต้องมี
- อักษรควบ ซึ่งในอักษรลาวปัจจุบันมี หุน (ຫ) และ หุม (ຫ) อยู่แล้ว แต่ในอักษรไทน้อยจะมีอักษรควบต่อไปนี้เพิ่มเติมด้วย:
 - ขน (ຂ) เช่น ຂອມ (ขนอม), ຂາງ (ขนาด)
 - ขุม (ຂ) เช่น ຂອງ (ขมอด)
 - คน (ດ) เช่น ດາ (คณา), ວັດະ (ลักคนะ), ດົງ (คินง)
 - คม (ດ) เช่น ດີ (คมี), ດັມ (คมัน)
 - ถน (ກ) เช่น ປາກາ (ปาถนา), ກອມ (ถนอม), ກົດ (ถนิต)
 - ถล (ກ) เช่น ແຮງ (ถลาง), ແຮງ (ถลึง)
 - สุน (ສ) เช่น ສັນ (สนั่น), ສົມ (สนม), ສຸກ (สนุก), ສະຫາ (เสน่หา), ສະ (เสนา)
 - สุม (ສ) เช่น ສົມ (ลัดสมี่), ສຸງ (สมุด), ສົມ (เสม็น [สุเมร]), ສົມ (เสมอ)
- อักษรเพ็อง มีการหีบยืมตัวเพ็องของอักษรธรรมมาใช้ เท่าที่พบได้แก่
 - ส เพ็อง (ງ) ใช้เป็นตัวสะกดแม่กด ปะปนกับ ด ธรรมดา เช่น ບຸ້ມີ (บัดนี้), ຮາງບຸ້ຊາ (ราดบุชชา)
 - ล เพ็องข้าง (ງ) เพิ่มเติมจาก ล เพ็องล่าง (ງ) ที่มีในอักษรลาวปัจจุบัน เช่น ບຸ້ (ปลี่ [ปี่]), ຫາງ (ทลาด), ບຸ້ທົບ (ปละทีบ), ຫາງ (ทลง [ทรง]), ຫາງ (ผลยง [เปลี้ยง]), ຫາງ (เสลา [เจลา]), ຫາງ (ทลาย [ทะลาย])

- บ เพ็อง (ບ) ใช้เป็นตัวสะกดแม่กบ ปะปนกับ บ ธรรมดา เช่น ຜູງ (ผาบ), ກັບ (กับ), ຈີ້ (ฮีบ), ຄວ (คาบ), ຂວ (ขาบ)
- ม เพ็อง (ມ) เช่น ມາມູ (นามมะ), ກູມາມ (กุ่มมาน), ອູມາ (อุมมา [อุ้มมา])
- พ เพ็อง (ພ) เช่น ທິພະໂສງ (ทิพพะโสด), ທິພະອາງ (ทิพพะอาด)
- ธิ เพ็อง (ັ) เช่น ອັທິຣາມ (อัทธิธาน [อธิษฐาน]), ສິທິເດັດ (สิทธิเต็ด [สิทธิเดช]), ສິທິ (สิทธิ [ฤทธิ])
- ด เพ็อง (ດ) เช่น ບຸ້ມີ (บัดนี้)
- จ เพ็อง (ຈ) เช่น ສະເຕັຈ (สะเต็จ [เสด็จ])
- ต เพ็อง (ຕ) เช่น ບຸຕະລາດ (บุตตะลาด [บุตรราช]), ຂັຕະລິກ (ขัตตะลิก [ขัตถฤษ])
- น เพ็อง (ນ) เช่น ຍົວ (อ्यान)
- สระออล่าง (ອ) เช่น ຫູ້ (ห่อ), ເພື່ (เพื่อ), ລູ້ (ล่อ)

๓ ทางเลือกของการกำหนดรหัส

การต่อเติมตารางรหัสยูนิโค้ดของอักษรลาวปัจจุบัน แยกประเด็นเป็น ๒ ส่วน คือสระออย และตัวควบ/ตัวเพ็อง

๓.๑ สระออย (ງ)

อาจกำหนดรหัสโดยอาศัยช่องตารางที่ยังว่างอยู่ในทั้งอักษรไทยและลาว ซึ่งช่องที่เหมาะสมน่าจะเป็น U+0EBE ซึ่งอยู่ถัดจาก ย เพ็อง (ງ U+0EBD) พอดี

๓.๒ อักษรควบและตัวเพ็อง

ความเป็นไปได้ในการรองรับอักษรควบและตัวเพ็อง ได้แก่

๑. ใช้วิราม โดยกำหนดช่องตารางที่ตรงกับพินทุของอักษรไทย ได้แก่ U+0EBA เพื่อใช้เป็นวิราม จากนั้นก็ใช้วิรามในการเชื่อมอักษรควบ (เช่น ค <OEBA> ม = คม) และใช้กำหนดให้อักขระถัดไปเป็นตัวเพ็อง (เช่น <OEBA> ฆ = ฌ)

เหตุผล: เป็นวิธีที่ยืดหยุ่นต่อการเพิ่มตัวควบและตัวเฟื่องที่อาจพบใหม่ โดยต้องการอักขระเพิ่มในตารางเพียงตัวเดียวเท่านั้น คือวิราม

วิธีการ: ฟอนต์กำหนดกฎ GSUB สำหรับแปลงลำดับอักขระที่เชื่อมด้วยวิราม ให้เป็นตัวควบหรือตัวเฟื่อง

ปัญหา: เป็นการเพิ่มกลไกของวิรามในบล็อกยูนิโค้ดอักษรลาว ซึ่งไม่เคยมีมาก่อน ทำให้ต้องใช้เวลากว่าจะย้ายระบบต่างๆ ให้มารองรับ

๒. กำหนดรหัสอักขระรายตัว โดยกำหนดรหัสอักขระสำหรับตัวควบหรือตัวเฟื่องเป็นรายตัว โดยอาจใช้ช่องตารางในช่วง U+0EE0 ถึง U+0EFF ที่ยังว่างอยู่ หรือจองช่วงรหัสต่อเติม (supplement) ช่วงใหม่ในยูนิโค้ดโดยเฉพาะ

เหตุผล: implement ได้ง่าย กลมกลืนกับตัวควบสองตัวที่มีอยู่แล้ว คือ หุน (໙) และ หุม (໘) และไม่ต้องเพิ่มความสามารถพิเศษจากโมเดลที่มีอยู่ของอักษรไทย-ลาวเพื่อมาจัดการวิราม

วิธีการ: เพิ่มรหัสอักขระสำหรับตัวควบและตัวเฟื่องในช่วง U+0EE0 ถึง U+0EFF ที่ยังว่างอยู่ หรือจองช่วงรหัสต่อเติม (supplement) ช่วงใหม่ในยูนิโค้ดโดยเฉพาะ

ปัญหา: ยังไม่ทราบตัวควบและตัวเฟื่องที่มีใช้ทั้งหมดในเอกสารโบราณ อาจพบเพิ่มได้อีกในอนาคต ซึ่งอาจทำให้ช่องตารางที่เตรียมไว้มีไม่พอ วิธีการนี้จึงอาจเป็นเพียงวิธีชั่วคราวเท่านั้น

๓. กำหนดรหัสตัวควบ และใช้วิรามสำหรับตัวเฟื่องอักษรธรรม โดยผสมทั้งสองวิธีการข้างต้น

เหตุผล: จำนวนตัวควบมีจำกัด ในขณะที่การยืมตัวเฟื่องอักษรธรรมเป็นไปอย่างอิสระ จึงมีความเป็นไปได้ที่จะจำกัดรายการตัวควบที่จะเพิ่ม แต่ปล่อยให้ใช้ตัวเฟื่องอักษรธรรมได้อย่างอิสระ นอกจากนี้ การกำหนดรหัสตัวควบยังทำให้สอดคล้องกับตัวควบ ໙ และ ໘ ที่มีอยู่ก่อน

ปัญหา: เช่นเดียวกับการใช้วิรามข้างต้น

๔. ใช้ ZWJ เชื่อมตัวควบ อักขระ ZERO WIDTH JOINER (U+200D) สามารถใช้เชื่อมอักขระสองตัวให้เป็นรูปเดียวได้ จึงสามารถใช้สร้างตัวควบที่น้อยได้ ส่วนตัวเฟื่องอักษรธรรมก็อาจใช้วิธีใดวิธีหนึ่งข้างต้น (ใช้วิรามหรือกำหนดรหัสรายตัว)

เหตุผล: ไม่ต้องกำหนดรหัสอักขระใหม่สำหรับตัวควบเลย สามารถใช้อักขระ
ยูนิโค้ดที่มีอยู่ได้ทันที

ปัญหา: ไม่สอดคล้องกับตัวควบ ๗ และ ๘ ที่มีอยู่ก่อน