



## การแปลงข้อความภาษาบาลีอักษรไทยเป็นสัทอักษร Grapheme to Phoneme Transcription for Pali-Thai

วินัย มาลีลัย<sup>1</sup> และ พุชชดี ศิริแสงตระกูล<sup>1\*</sup>

### บทคัดย่อ

บทความนี้นำเสนอการแปลงภาษาบาลีอักษรไทยเป็นสัทอักษร โดยใช้พจนานุกรม (dictionary) ร่วมกับกฎบาลีไวยากรณ์ (Pali grammar) และภาษาศาสตร์ (linguistic) และได้ใช้วิธีการตัดคำแบบเลือกคำที่ยาวที่สุด (longest matching algorithm) เพื่อใช้ในการค้นหาความหมายของคำศัพท์ ในงานวิจัยนี้ผู้วิจัยได้สร้างพจนานุกรมภาษาบาลี-ไทย ประกอบด้วยคำศัพท์บาลี 14,323 คำ โดยอ้างอิงคำศัพท์บาลีจากหนังสือพจนานุกรมภาษาบาลี-ไทยของพระมหาไพโรจน์ พจนานุกรมที่สร้างขึ้นประกอบด้วยคำศัพท์ภาษาบาลีอักษรไทย ความหมายประเภทของคำศัพท์ คำในรูปภาษาไทย คำในรูปอักษรโรมันและคำในรูปสัทอักษร โดยจัดเก็บไว้ในระบบฐานข้อมูล และในงานวิจัยนี้ได้นำเสนอกฎในการแปลงภาษาบาลีเป็นสัทอักษรเพื่อใช้ในการแปลงคำที่ไม่ปรากฏในพจนานุกรมเพื่อทดสอบความถูกต้องของระบบ ผู้วิจัยได้ทำการทดสอบระบบที่พัฒนาโดยการแปลงเอกสารที่เขียนด้วยภาษาบาลีอักษรไทย ซึ่งประกอบด้วย บทสวดมนต์ 99 บท พุทธศาสนสุภาษิต 181 สุภาษิต และวินย. มหาวิภังโก 1 จากพระวินยปิฎก เล่มที่ 1 ของพระไตรปิฎกภาษาบาลีฉบับสยามรัฐ ผลการทดสอบความถูกต้องของการแปลงคิดเป็นร้อยละ 98.75

<sup>1</sup> ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น อำเภอเมือง จังหวัดขอนแก่น 40002

\*Correspondent Author, E-mail: pusadee@kku.ac.th

## ABSTRACT

This paper presents a method for Pali phoneme transcription using a hybrid of dictionary and linguistic rules. In the study, we constructed a Pali-Thai dictionary, which gathered Pali words from the dictionary book of Phramahapiroch. The dictionary consists of almost 14,323 Pali words. The dictionary structure contains Pali words, meanings, word type, corresponding Thai words, romanize words, as well as phoneme. The input of the developed system is Pali language in paragraph level. The input sentence was segmented into sequence of words using the longest matching algorithm in order to look for Pali word and phoneme from the dictionary. Words that do not appear in the dictionary are transcribed using the Pali grammar and linguistic rules. In order to evaluate the efficiency of the proposed system, the developed system was used to transcribe Pali documents. The testing documents are 99 prayers, 181 proverbs and Tipitaka Volume 1 of siamrath. The transcription result shows that the accuracy the transcription was 98.75%.

**คำสำคัญ:** ภาษาบาลี สัทอักษร การตัดคำแบบเลือกคำที่ยาวที่สุด การแปลงข้อความ

**Key Words:** Pali, Phoneme, Longest matching, Transcription

## บทนำ

ประเทศไทยได้รับเอาศาสนาพุทธเป็นศาสนาประจำชาติ ซึ่งศาสนาพุทธจะใช้ภาษาบาลีในการเผยแผ่ศาสนา ภาษาบาลีเป็นภาษาที่ไม่มีการลงตัวอักษรโดยตรง เมื่อประเทศใดนำภาษาบาลีมาใช้จึงมีการจารึกภาษาบาลีด้วยอักษรของประเทศตนเอง (ประวัติความเป็นมาและพัฒนาการของภาษาบาลี, 2555) เช่น อักษรพราหมี อักษรเทวนาครี อักษรล้านนา อักษรขอม อักษรไทย เป็นต้น คำสอนในทางพุทธศาสนาได้ถูกรวบรวมไว้เป็นหมวดหมู่เรียกว่าพระไตรปิฎก มี 3 หมวดคือ พระวินัยปิฎก พระสุตตันตปิฎก และพระอภิธรรมปิฎก โดยถูกบันทึกไว้เป็นภาษาบาลีทั้งสิ้น

ปัจจุบันการเรียนการสอนภาษาบาลีอักษรไทยส่วนใหญ่มีการเรียนการสอนเฉพาะผู้ที่บวชเรียน รวมทั้งการสนับสนุนหรือการให้ความรู้เกี่ยวกับภาษาบาลีที่ถูกต้องยังไม่มีแพร่หลายในกลุ่มบุคคล

ทั่วไป จึงทำให้บุคคลทั่วไปที่มีความสนใจที่จะศึกษาคำสอนทางพุทธศาสนาประสบปัญหาในการอ่านภาษาบาลี เนื่องจากภาษาบาลีมีการเขียนและการอ่านแตกต่างจากภาษาไทยรวมถึงมีการใช้อักษรพิเศษคือ พินทุและนิคหิตเข้ามามีส่วนร่วมในการอ่าน จึงทำให้การอ่านภาษาบาลีมีความซับซ้อนยิ่งขึ้น เช่น เอตมมงคฺลุมตฺตมฺ อ่านว่า เอตัมมังคะละมุตตะมัง เป็นต้น ปัจจุบันมีงานวิจัยหลายงานได้นำเสนอวิธีการแปลงภาษาของประเทศตนเองให้อยู่ในรูปแบบสัทอักษร เนื่องจากสัทอักษรเป็นรูปแบบมาตรฐานที่ใช้ในการออกเสียงเพื่อใช้ในการสื่อสารของภาษาทุกภาษาในโลก (จรัลวิไล, 2555) รวมถึงมีการใช้สัทอักษรในการประมวลผลภาษาธรรมชาติ จากงานวิจัยที่ผ่านมาได้มีการนำเสนอวิธีการในการแปลงข้อความเป็นสัทอักษรโดยใช้วิธีการต่าง ๆ เช่น การใช้กฎภาษาศาสตร์ร่วมกับพจนานุกรม (นงคฺนุชและพุทธศติ, 2552) วิธีการของ

ต้นไม้ตัดสินใจที่ใช้ในการแยกโครงสร้างภาษา (Ordean et al., 2009) วิธีการถอดโครงสร้างตามหลักไวยากรณ์ของภาษา (บวรรัตน์และพุทธศรี, 2549) เป็นต้น ซึ่งก็ยิ่งประสบกับปัญหาของความถูกต้องในการแปลง เนื่องจากโครงสร้างของภาษามีความแตกต่างกัน

งานวิจัยนี้จึงได้นำเสนอวิธีการแปลงภาษาบาลีอักษรไทยเป็นสัทอักษร โดยใช้พจนานุกรมร่วมกับกฎบาลีไวยากรณ์และกฎภาษาศาสตร์ เพื่อให้ภาษาบาลีเป็นรูปแบบมาตรฐานกลางที่เหมาะสมสำหรับใช้ในการศึกษาหลักธรรมคำสอนทางพุทธศาสนาของคนทั่วโลก ซึ่งเมื่อประเทศใดนำภาษาบาลีในรูปแบบสัทอักษรไปใช้ก็เพียงแค่ทำการเทียบอักษรของประเทศตนเองกับสัทอักษรก็สามารถทำการศึกษาภาษาบาลีได้อย่างถูกต้อง

## ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 1. ภาษาบาลี

พระมหานิยม (2549) ภาษาบาลีอักษรไทยประกอบด้วยพยัญชนะ 33 ตัว สระ 8 ตัว อักษรพิเศษ 2 ตัว ดังปรากฏในตารางที่ 1 และ 2 ซึ่งภาษาบาลีอักษรไทยมีหลักการอ่านดังนี้

1. พยัญชนะที่ไม่มีสระอาศัยอยู่ให้อ่านออกเสียงสระ อะ เช่น พล อ่านว่า พละ

2. พยัญชนะที่อยู่ข้างหน้าพยัญชนะตัวที่มีพินทุ ไม่ประกอบด้วยสระใด ๆ ให้อ่านออกเสียงแบบเดียวกับไม้หันอากาศในภาษาไทย เช่น ฦนเต อ่านว่า ฦนเต

3. พยัญชนะที่มีสระอาศัยอยู่และพยัญชนะตัวถัดไปมีพินทุอยู่ได้ พยัญชนะที่มีพินทุอยู่ได้จะทำหน้าที่เป็นตัวสะกดเท่านั้น เช่น พุท อ่านว่า พุท

4. พยัญชนะตัวใดที่มีนิคหิตอยู่ข้างบน ให้อ่านออกเสียงเป็นแม่กง คือใช้ตัว ‘ง’ เป็นตัวสะกด เช่น อรหฺ อ่านว่า อะระหัง

5. นิคหิตเมื่อมาประกอบกับสระ อะ อิ อุ คือ อํ อี อู ออกเสียงเป็น อัง อิง อุง

6. พยัญชนะ 4 ตัว คือ ย ร ล ว อยู่หลังพยัญชนะตัวอื่น และพยัญชนะข้างหน้าตนที่มีพินทุอยู่ได้ ให้ออกเสียงควบกับพยัญชนะตัวหน้า โดยที่พยัญชนะที่มีพินทุอยู่ได้ไม่ต้องอ่านออกเสียงสระอะ เช่น ชิวหา อ่านว่า ชิวหา

7. พยัญชนะตัวใดที่มีพินทุอยู่ได้ ซึ่งเป็นตัวแรกของคำนั้น ๆ และพยัญชนะถัดไปไม่ใช่พยัญชนะ 4 ตัว คือ ย ร ล ว พยัญชนะที่มีพินทุอยู่ได้ ให้อ่านออกเสียงสระอะ เช่น สมฺ อ่านว่า สะมุ

**ตารางที่ 1** รูปแบบสัทอักษรของพยัญชนะภาษาบาลี (จรัลวิไล, 2555)

พยัญชนะ บาลีอักษรไทย	สัทอักษร	
	พยัญชนะต้น	พยัญชนะท้าย
ก	/k/	/k/
ข ค ฦ	/k <sup>h</sup> /	/k/
ง	/ŋ/	/ŋ/
จ	/c/	/t/
ฉ ช ฦ	/c <sup>h</sup> /	/t/
ญ	/j/	/n/
ฎ ต	/t/	/t/
ฐ ฑ ฒ ณ ฑ ฐ	/t <sup>h</sup> /	/t/
ณ น	/n/	/n/
ป	/p/	/p/
ผ พ ภ	/p <sup>h</sup> /	/p/
ม	/m/	/m/
ย	/j/	/j/
ร	/r/	/n/
ล พ	/l/	/n/
ว	/w/	/w/
ส	/s/	/t/
ห	/h/	
อ	/ʔ/	

**ตารางที่ 1** รูปแบบสัทอักษรของพยัญชนะภาษาบาลี (จรัลวิไล, 2555) (ต่อ)

อักษรควบกล้ำ		อักษรควบกล้ำ	
บาลีอักษรไทย	สัทอักษร	บาลีอักษรไทย	สัทอักษร
กร	/kr/	ปร	/pr/
กล	/kl/	ปล	/pl/
กว	/kw/	พร	/p <sup>h</sup> r/
ขร คร	/k <sup>h</sup> r/	พล ผล	/p <sup>h</sup> /
ขล คล	/k <sup>h</sup> l/	ตร	/tr/
ขว คว	/k <sup>h</sup> w/		

**ตารางที่ 2** รูปแบบสัทอักษรของสระและวรรณยุกต์ ภาษาบาลี (จรัลวิไล, 2555)

สระบาลี อักษรไทย	สัทอักษร	อักขระพิเศษ	สัทอักษร
อะ	/a/	. (พินทุ)	
อา	/a:/	° (นิคหิต)	/aŋ/
อิ	/i/	เสียงวรรณยุกต์	สัทอักษร
อี	/i:/	สามัญ	
อุ	/u/	เอก	`
อุ	/u:/	โท	^
เอ	/e:/	ตรี	'
โอ	/o:/	จัตวา	˘

## 2. งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องมีผู้ทำวิจัย การแบ่งพยางค์ การตัดคำ และการแปลงข้อความ เป็นสัทอักษร ดังนี้

ยีนและวิวรรธน์ (2529) นำเสนอการแบ่งพยางค์ด้วยพจนานุกรม โดยใช้วิธีการเทียบคำที่ยาวที่สุดในการตัดแบ่งพยางค์ และในกรณีที่ไม่พบคำในพจนานุกรม ผู้วิจัยได้มีการนำกฎไวยากรณ์จำนวน 18 กฎ เข้ามาช่วยในการแบ่งพยางค์ ความถูกต้องอยู่ที่ร้อยละ 99

นงค์นุชและพุทธิชัย (2552) งานวิจัยนี้ได้ นำเสนอการแปลงอักษรธรรมอีสานเป็นสัทอักษร โดยผู้วิจัยได้ใช้วิธีการตัดคำแบบเลือกคำยาวที่สุด (longest

matching) ในการสืบค้นคำในพจนานุกรม ในกรณีที่ไม่พบคำในพจนานุกรม ผู้วิจัยจะใช้กฎตามโครงสร้างคำของอักษรธรรมอีสานร่วมกับกฎตามหลักภาษาศาสตร์ในการแปลงเป็นสัทอักษร ความถูกต้องอยู่ที่ร้อยละ 81.92

บวรรัตน์และพุทธิชัย (2549) งานวิจัยนี้ได้ นำเสนอวิธีการแปลงรูปคำภาษาไทยเป็นสัทอักษร โดยใช้กฎภาษาศาสตร์และพจนานุกรมภาษาไทย ซึ่งจัดเก็บข้อมูลในโครงสร้างแบบทรีและได้นำวิธีการตัดคำแบบเลือกคำยาวที่สุดมาใช้ในการตัดคำและแบ่งพยางค์ ความถูกต้องอยู่ที่ร้อยละ 90.09

Ordean et al. (2009) นำเสนอวิธีการปรับปรุงกฎการแปลงสัทอักษรสำหรับภาษาโรมาเนีย โดยใช้กฎไวยากรณ์ของภาษาโรมาเนียร่วมกับพจนานุกรมภาษาโรมาเนีย (DOOM II dictionary) และใช้ต้นไม้ตัดสินใจสำหรับกฎยกเว้นของคำที่ไม่มีในหลักไวยากรณ์ โดยใช้อัลกอริทึมดัชนีจีนิ (Gini Index) เป็นมาตรการ ที่จะประเมินจุดแยกแต่ละโหนดที่ต้องการยกเว้น ซึ่งผลของการวิจัยนี้ได้ความถูกต้องอยู่ที่ร้อยละ 99

## วิธีการดำเนินงานวิจัย

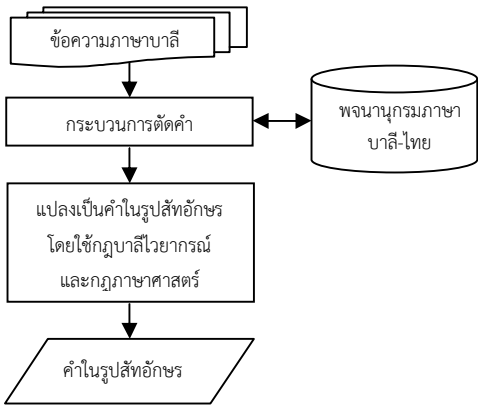
งานวิจัยนี้ แบ่งการทำงานออกเป็น 3 ขั้นตอน คือ 1) การออกแบบโครงสร้างและสร้างพจนานุกรมภาษาบาลี 2) กระบวนการตัดคำ 3) กระบวนการแปลงข้อความภาษาบาลีอักษรไทยเป็นสัทอักษร ดังแสดงในรูปที่ 1

### 1. สร้างพจนานุกรมภาษาบาลี

การสร้างฐานข้อมูลพจนานุกรมในงานวิจัยนี้ ผู้วิจัยได้ใช้ระบบจัดการฐานข้อมูลไมโครซอฟท์แอคเซส (Microsoft access) เนื่องจากระบบจัดการฐานข้อมูลไมโครซอฟท์แอคเซสสามารถจัดเก็บตัวอักขระของสัทอักษรได้ครบทุกตัว จึงทำให้การประมวลผลของระบบ

ทำได้เร็วขึ้นเนื่องจากไม่ต้องทำการแทนค่าหรือแปลงค่าใด ๆ รวมทั้งระบบจัดการฐานข้อมูลไมโครซอฟท์แอคเซสไม่จำเป็นต้องทำการตั้งค่าใดๆก็สามารถนำไปใช้งานได้ (นันทน์, 2555) ซึ่งคำศัพท์ที่นำมาบรรจุลงในพจนานุกรมเป็นคำศัพท์ที่นำมาจากพจนานุกรมนามศัพท์ภาษาบาลี จำนวน 14,323 คำ (พระมหาไพโรจน์, 2553) ประกอบด้วย คำศัพท์ภาษาบาลี ความหมายของคำศัพท์ และประเภทของคำศัพท์

ในการวิจัยนี้ผู้วิจัยได้ทำการออกแบบฐานข้อมูลพจนานุกรมเพื่อให้รองรับการประมวลผลภาษาธรรมชาติที่เกี่ยวข้องกับภาษาบาลีต่อไปในอนาคต โครงสร้างพจนานุกรมที่สร้างขึ้นประกอบด้วย คำศัพท์ภาษาบาลี ความหมายของคำศัพท์ ประเภทของคำศัพท์ คำเขียนในรูปภาษาไทย คำในรูปอักษรโรมันและสัทอักษร โครงสร้างพจนานุกรมและตัวอย่างการจัดเก็บคำศัพท์ในพจนานุกรม ดังแสดงในตารางที่ 3 และ 4 ตามลำดับ



รูปที่ 1 ภาพรวมของระบบการแปลงภาษาบาลีอักษรไทยเป็นสัทอักษร

2. การตัดคำภาษาบาลีไทย (Pali word segmentation)

งานวิจัยนี้ใช้การตัดคำแบบวิธีการเทียบคำที่ยาวที่สุด (longest matching) ซึ่งเป็นวิธีการตัดคำที่

ต้องใช้ร่วมกับพจนานุกรมเพื่อค้นหาค่าและความหมายของคำในประโยค โดยมีกระบวนการทำงานดังแสดงในรูปที่ 2

```

C = {c1,c2,...,cn};      # ชุดของตัวอักษรในประโยค
D = {d1,d2,...,dm};      # ชุดของคำในพจนานุกรม
W = {w1,w2,...,wn};      # ชุดผลลัพธ์ของคำ
N = length(C);           # จำนวนอักขระในประโยค
i = 1;                   # ตำแหน่งของคำที่แสดงออกมา
SP = 1;                 # มาร์คตำแหน่งเริ่มต้น
LP = N;                 # มาร์คตำแหน่งสุดท้าย
TN = N;                 # จำนวนอักขระในประโยคที่เริ่มนับใหม่
DO WHILE (TN > 0)
    IF (C[SP to LP ]=D) THEN
        W[i]=C[SP to LP];
        i = i+1;
        SP = LP+1;
        C = C[Csp to cn];
        LP = N;      TN = length(C);
    ELSE
        IF (LP = SP) THEN
            W[i] = C[csp];
            i = i+1;
            SP = LP+1;
            C = C[Csp to cn];
            LP = N;
            TN = length(C);
        ELSE
            LP = TN-1;
        END IF
    END IF
LOOP
    
```

รูปที่ 2 กระบวนการตัดคำโดยการเทียบคำที่ยาวที่สุด

ตารางที่ 3 โครงสร้างพจนานุกรม

NO	Entity	Type	Caption
1	IdNo (PK)	AutoNumber	รหัสอัตโนมัติ
2	Pali	Text (100)	คำศัพท์
3	Mean	Text (255)	ความหมาย
4	Type	Text (10)	ประเภทคำศัพท์
5	Thai	Text (150)	คำภาษาไทย
6	Roman	Text (180)	โรมันในเซชั่น
7	Phonetic	Text (180)	สัทอักษร



ตารางที่ 5 กฎการแปลงภาษาบาลีอักษรไทยเป็นสัทอักษร

กฎบาลีไวยากรณ์	โครงสร้างกฎ	ตัวอย่างคำในกฎ	สัทอักษร
1) [C]ะ[T]	CVT	อ	?à
2) [C]ั[T][S]	CVTS	อัม	t <sup>h</sup> am
3) [เ โ] [C][T][S]	VCTS	เทพ	t <sup>h</sup> è:p
4) [C][า ิ ี ึ ุ ู][T][S]	CVTS	มารุ	ma:n
5) [C]ั[T][S]	CVTS	อัม	?an
6) [C]ิ[T][S]	CVTS	อิม	?iη
7) [C]ุ[T][S]	CVTS	อุม	?uη
8) [C][ย ร ล ว]	CC	พร	p <sup>h</sup> r
9) [C][T][N]	CVT	สมุ	sà mú
10) [เ โ] [H][T][L]	VCTS	เสย	sè:j
11) [เ โ] [M][T][L]	VCTS	เกย	ke:j
12) [เ โ] [L][T][L]	VCTS	เนย	ne:j
13) [H][ะ ั ิ ี ึ ุ ู][T][L]	CVTS	สาย	sá:j
14) [M][ะ ั ิ ี ึ ุ ู][T][L]	CVTS	กาย	ka:j
15) [L][ะ ั ิ ี ึ ุ ู][T][L]	CVTS	นาย	na:j
16) [เ โ] [H][T]	VCT	เส	sé:
17) [เ โ] [M][T]	VCT	เก	ke:
18) [เ โ] [L][T]	VCT	เน	ne:
19) [H][า ิ ี ึ ุ ู][T]	CVT	สา	sá:
20) [M][า ิ ี ึ ุ ู][T]	CVT	กา	ka:
21) [L][า ิ ี ึ ุ ู][T]	CVT	นา	na:
22) [เ โ] [H][T][D]	VCTS	เสก	sè:k
23) [เ โ] [M][T][D]	VCTS	เกก	kè:k
24) [เ โ] [L][T][D]	VCTS	เนก	nè:k
25) [H][ะ ั ิ ี ึ ุ ู][T][D]	CVTS	สาก	sà:k
26) [M][ะ ั ิ ี ึ ุ ู][T][D]	CVTS	กา	kà:k
27) [L][ะ ั ิ ี ึ ุ ู][T][D]	CVTS	นา	nâ:k
28) [L][ะ ิ ี ึ ุ ู][T][D]	CVTS	นิก	ník
29) [H][ะ ิ ี ึ ุ ู][T]	CVT	สิ	sì
30) [M][ะ ิ ี ึ ุ ู][T]	CVT	กิ	kì
31) [L][ะ ิ ี ึ ุ ู][T]	CVT	นิ	nì

ตารางที่ 6 กฎการปรับเปลี่ยนโครงสร้าง เช่น ธม /ธัม/

ภาษาบาลีอักษรไทย	ธ		ม
กฎบาลีไวยากรณ์	[C]	~	[T] [S]
โครงสร้างของกฎ	C	V	T S
รูปภาษาไทย	ธ	~	ม
รูปสัทอักษร	t <sup>h</sup>	a	m

### การทดลองและผลการทดลอง

เพื่อทดสอบประสิทธิภาพของระบบที่พัฒนาขึ้นผู้วิจัยได้ใช้เอกสารที่เขียนด้วยภาษาบาลีอักษรไทย ซึ่งประกอบด้วย บทสวดมนต์ 99 บท พุทธศาสนสุภาษิต 181 สุภาษิต และวินย. มหาวิทยาลัยโค 1 จาก พระวินัยปิฎก เล่มที่ 1 ของพระไตรปิฎกภาษาบาลีฉบับสยามรัฐ ทำการแปลงเป็นสัทอักษรด้วยระบบที่พัฒนาขึ้น และทำการวัดประสิทธิภาพ ของการแปลงภาษาบาลีเป็นสัทอักษรโดยใช้ตัวชี้วัดร้อยละความถูกต้อง คือค่าที่แปลงถูกต้องต่อจำนวนทั้งหมดคิดเป็นเปอร์เซ็นต์แสดงได้ดังสมการ ดังนี้

$$\text{ประสิทธิภาพ (\%)} = \frac{X}{N} \times 100$$

โดยที่ X คือจำนวนค่าที่แปลงถูกต้อง

N คือค่าทั้งหมด

ผลการทดสอบประสิทธิภาพของการแปลงข้อความภาษาบาลีอักษรไทยเป็นสัทอักษร ดังแสดงในตารางที่ 7

ตารางที่ 7 ผลการแปลงข้อความภาษาบาลีอักษรไทย

ประเภทเอกสาร	จำนวนค่า	แปลงถูก	คิดเป็น %
บทสวดมนต์ 99 บท	2,245	2,187	97.41
พุทธศาสนสุภาษิต 181 สุภาษิต	464	464	100.00
วินย. มหาวิทยาลัยโค 1 หน้าที่ 1 ถึง 100	7,458	7,373	98.86
รวม	10,167	10,024	98.75

ในการตรวจสอบความถูกต้องของการแปลงข้อความภาษาบาลีเป็นคำในรูปภาษาไทย ผู้วิจัยได้นำ

คำในรูปภาษาไทยที่แปลงได้ ไปเทียบเคียงความถูกต้องกับเอกสารบททำวัตร สวดมนต์ (กองพุทธศาสนศึกษา, 2553) และบทเจริญพระพุทธรูป สวดมนต์ข้ามปี (กองพุทธศาสนศึกษา, 2554) และในการแปลงภาษาบาลีเป็นสัทอักษร ผู้วิจัยได้ใช้วิธีการปรัวรรตตัวอักษรจากอักษรไทยเป็นสัทอักษร ตามหลักการเทียบเคียงอักษรไทยเป็นสัทอักษรสากล (จรัลวิไล, 2555) ดังแสดงในตารางที่ 1 และตารางที่ 2

ตัวอย่างข้อความภาษาบาลีบางส่วนใน พระวินัยปิฎก เล่มที่ 1 หน้าที่ 1 ที่ใช้ทดสอบและผลลัพธ์รูปสัทอักษร

### ภาษาบาลีอักษรไทย

เตน สมเณ พุโธ ภควา เวรณุชา ย วิหระติ นเพ  
รุจุจินนทมูล มหตา ภิกขุสงเณน สทธิ ปญจมตเตหิ  
ภิกขุสเดหิฯ

อสโสสิ โข เวรณุโข พราหมโณ สมโณ ชลฺ โภ  
โคตโม สกยปุโต สกยกุล ปัพพะชิต เวรณุชา ย วิหระติ  
นเพรุจุจินนทมูล มหตา ภิกขุสงเณน สทธิ ปญจมตเตหิ  
ติ โข ปน ภวนติ โคตมิ เอว กल्याโณ กิตติสทโท อพ  
ภุคคะโต อติปิ โส ภควา

### คำในรูปภาษาไทย

เตนะ สมเณนะ พุโธ ภคควา เวรณุชา ยัง  
วิหระระติ นะเพรุจุจินนทมูล มะหะตะตา ภิกขุสงเณนะ  
สทธิ ปญจจะมัตเตหิ ภิกขุสเดหิฯ

อัสโสสิ โข เวรณุโข พราหมโณ สมโณ ชลฺ  
โภ โคตะโม สักยะปุโต สักยะกุล ปัพพะชิต เวรณุชา  
ยัง วิหระระติ นะเพรุจุจินนทมูล มะหะตะตา ภิกขุสงเ  
นะ สทธิ ปญจจะมัตเตหิ ตัง โข ปะนะ ภาวันตัง โค  
ตะมัง เอวัง กัลยาโณ กิตติสทโท อพภุคคะโต อติปิ โส  
ภคควา

### คำในรูปสัทอักษร

te:ná sámáje:ná p<sup>h</sup>útt<sup>h</sup>o: p<sup>h</sup>ák<sup>h</sup>áwa:  
we:ranc<sup>h</sup>a:jaṅ wíhàrátí nále:rúpucimant<sup>h</sup>ámu:le:



máhàta: p<sup>h</sup>ikk<sup>h</sup>ũṅṣaṅk<sup>h</sup>e:ná sàtt<sup>h</sup>iṅ pancámátte:hi  
p<sup>h</sup>ikk<sup>h</sup>ũsàte:hi

ʔàtsò:si k<sup>h</sup>õ: we:ranc<sup>h</sup>o: p<sup>h</sup>ra:mno:

sámáno: k<sup>h</sup>àlú p<sup>h</sup>o: k<sup>h</sup>o:tàmo: sàkjápùtto:

sàkjákùla: pàpp<sup>h</sup>á-c<sup>h</sup>ító: we:ranc<sup>h</sup>a:jaṅ wihàrāti

nále:rúpucimant<sup>h</sup>ámu:le: mähàta:

p<sup>h</sup>ikk<sup>h</sup>ũsāṅk<sup>h</sup>e:ná sàtt<sup>h</sup>iṅ pancámátte:hi taṅ k<sup>h</sup>õ:

páná p<sup>h</sup>áwantāṅ k<sup>h</sup>o:tāmaṅ ʔe:wāṅ kanja:no:

kittisatt<sup>h</sup>o: ʔāpp<sup>h</sup>ũkk<sup>h</sup>áto: ʔitipi sò: p<sup>h</sup>ák<sup>h</sup>áwa:

### อภิปรายและสรุปผลการวิจัย

งานวิจัยนี้พัฒนาขึ้นเพื่อทำการแปลงข้อความภาษาบาลีอักษรไทยเป็นสัทอักษร โดยใช้พจนานุกรมร่วมกับกฎบาลีไวยากรณ์และกฎภาษาศาสตร์ ในการพัฒนาระบบผู้วิจัยได้ทำการสร้างพจนานุกรมภาษาบาลีขึ้น และในการประมวลผลข้อความที่นำเข้าจะถูกตัดคำด้วยวิธีการเทียบคำที่ยาวที่สุด จากผลการทดสอบเพื่อประเมินประสิทธิภาพความถูกต้องของระบบมีความถูกต้องที่ร้อยละ 98.75 ผลการวิเคราะห์พบว่าข้อผิดพลาดส่วนใหญ่เกิดจากการแปลงคำควบกล้ำ เช่น คำว่า “ทริ” หากอ่านตามกฎบาลีไวยากรณ์จะอ่านว่า “ทะริ” แต่เมื่ออ่านแบบคำควบกล้ำจะอ่านว่า “ทริ” และคำว่า “ทวิ” การอ่านที่ต้องจะอ่านว่า ทวิ ซึ่งจะเห็นว่าพยัญชนะตัวที่สองของทั้งสองคำคือ ร และว ซึ่งเป็นพยัญชนะอักษรควบกล้ำในภาษาบาลี(ย ร ล ว) แต่การอ่านออกเสียงมีความแตกต่างกัน จึงทำให้ผลของการแปลงคำควบกล้ำมีข้อผิดพลาด ดังนั้นเพื่อให้ระบบมีความสมบูรณ์ครบถ้วนเหมาะสำหรับการศึกษภาษาบาลี ควรทำการเพิ่มโครงสร้างกฎเพื่อให้สามารถแปลงคำภาษาบาลีที่เป็นคำควบกล้ำได้อย่างถูกต้อง รวมถึงรวบรวมคำควบกล้ำในภาษาบาลีให้ครบทุกคำหรือให้ได้มากที่สุด เพื่อเป็นการกำหนดคุณสมบัติเฉพาะให้กับคำเหล่านั้น แนวทางในการพัฒนาต่อไป เพื่อให้เกิดประโยชน์มากขึ้น ควรพัฒนาใน

ส่วนของ การแปลงข้อความและประโยคบาลีเป็นภาษาไทย เพื่อให้ผู้ที่ศึกษาสามารถเข้าใจถึงแก่นมากขึ้น

### เอกสารอ้างอิง

- กองพุทธศาสนศึกษา, สำนักงานพุทธศาสนาแห่งชาติ. (2553). บททำวัตรสวดมนต์(แปล). กรุงเทพฯ: สำนักงานพุทธศาสนาแห่งชาติ.
- กองพุทธศาสนศึกษา, สำนักงานพุทธศาสนาแห่งชาติ. (2554). บทเจริญพระพุทธรูป สวดมนต์ข้ามปี. กรุงเทพฯ: สำนักงานพุทธศาสนาแห่งชาติ.
- จรัลวีโล จรุงโรจน์. (2555). ภาษาศาสตร์เบื้องต้น. พิมพ์ครั้งที่ 5. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- นงศุข ไพบูลย์ และพุชตี ศิริแสงตระกูล. (2552). การแปลงอักษรธรรมอีสานเป็นสัทอักษร. The 13<sup>th</sup> National Computer Science and Engineering Conference, กรุงเทพฯ, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.
- นันทินี แวงโสภณ. (2555). คู่มือ Access 2007 ฉบับสมบูรณ์ สำหรับปี 2012-2013. กรุงเทพฯ: โปรวิชั่น.
- บรรรัตน์ ศรีมาน และพุชตี ศิริแสงตระกูล. (2549). การแปลงคำภาษาไทยเป็นสัทอักษร โดยใช้กฎทางภาษาศาสตร์และอัลกอริทึมแบบเลือกค้ายาวที่สุด. วารสารวิจัย มข. (ฉบับบัณฑิตศึกษา) 6(2): 58-67.
- ประวัติความเป็นมาและพัฒนาการของภาษาบาลี. (2555). ค้นเมื่อ 15 ธันวาคม 2555, จาก <http://www.buddhabucha.net/history-of-pali-language>
- พระมหานิยม อุตตโม (ป.ธ.7). (2549). หลักสูตรย่อบาลีไวยากรณ์ฉบับสมบูรณ์. พิมพ์ครั้งที่ 4. กรุงเทพฯ: เลียงเชียง.
- พระมหาไพโรจน์ ปณฺญาวชิโร (ป.ธ.9). (2553). พจนานุกรมบาลี-ไทย. กรุงเทพฯ: โรงพิมพ์มหาจุฬาลงกรณราชวิทยาลัย.
- ยี่น ภู่วรรณ และวิวรรธน์ อิ่มอารมณ. (2529). การแบ่งแยกพยางค์ไทยด้วยดิคชันนารี. เอกสารประกอบการประชุมวิชาการวิศวกรรมไฟฟ้า สถาบันอุดมศึกษาแห่งประเทศไทย ครั้งที่ 9, มหาวิทยาลัยขอนแก่น.

สำนักงาน กศน. สำนักงานปลัดกระทรวงศึกษาธิการ. (2554).  
หนังสือเรียนสาระความรู้พื้นฐาน รายวิชา หลัก  
ภาษาไทย (พท 33002) ระดับมัธยมศึกษาตอน  
ปลาย. มหาสารคาม: สำนักงาน กศน.

Ordean, M.A., Saupe, A., Ordean, M., Duma, M., and  
Silaghi, G.C. (2009). Enhanced Rule-based

Phonetic Transcription for the Romanian  
Language. International Symposium on  
Symbolic and Numeric Algorithms for  
Scientific Computing. Romania: Timisoara.  
pp. 401-406.

