Thai spelling reform for data processing

UDOM Warotamasikkhadit
Ramkhamhaeng University

1. Introduction

Those who know Thai know that it is impossible to tell whether a vowel a, o, 
ɔ or null is to be inserted after the first consonant when two consonants stand next
each other. Very often even a native Thai speaker, who knows very well the
meaning of the words he reads, makes wrong syllabification cuts in words like
ขั้นมีถวายสร้อย, ขั้น-มา-บก-ระ, khöm flattowel, ขั้น-มา-บก-ระ
khöm Flallandp, ขั้น-มา-บก-ระ, khöm Chattlandp, ขั้น-มา-บก-ระ
khöm Chattlandp, or ขั้น-มา-บก-ระ, khöm Chattlandp where only the
last reading is acceptable. It, thus, is very difficult for a foreigner or a machine to
syllabify the written words in Thai correctly. This paper aims to present the problems
of syllabification, romanization, and data processing of Thai words and to propose
the methods of correct syllabification by man or machine.

2. Syllabification problems

2.1 It is difficult to tell whether the covert vowel is an a, o, or ɔ as in
สังฆินทร์ sānāk where the first covert vowel is ɔ and the second covert vowel is a,
และ pāmōn where the first covert vowel is a and the second covert vowel is
ɔ, ต่ำ tālp where the first covert vowel is a and the second covert vowel is o.

2.2 A ′, a silence marker, usually silences a character it is placed over, but there are
a number of Thai words where a silence marker silences up to three characters. It is
almost impossible to tell which characters are to be silent, for example: ไห่ wít
where only one character, ɔ, is silenced, ไห่ phoo where two characters -- a
consonant ɔ and a vowel -- are silenced, แหม่น kaan where two consonants, ɔ
and ɔ, are silenced, ล้า ek where two consonants, ɔ and ɔ, are silenced,
and ล้า ek where three consonants, ɔ, ɔ, and ɔ, are silenced.

2.3 It is impossible to tell whether ɔ or ɔ is a syllable final or a cluster of an initial,
for example: กระแส can be read as konmānaa or krommānaa, กระแส can be read as

TAI STUDIES PROCEEDINGS 405-410
konlámóp, klommaphá, kàlámóp, or kàlomphá. Those who know Thai will recognize that only krommánaa and kàlomphá are acceptable.

2.4 It is difficult to tell whether a character is a syllable final or an initial of the next syllable. นิพน can be read as kaanwêek or kaarâwêek. Those who know Thai will recognize that the second reading is acceptable and kaanwêek does not exist in Thai.

2.5 It is hard to tell whether แ is an initial แ, a second member of a cluster, a vowel แ, or a final แ. Similarly it is hard to tell whether ้ is an initial ้ or a vowel ้. ว can be read as khâwǒnkhâwâay, khwǒnkhwâay, or khûankhwâay. The correct pronunciation of this word according to the Dictionary of the Royal Institute is khwǒnkhwâay, but there are a number of people who pronounce it as khûankhwâay. คว can be read as khraaw?ôt or khraawšôt, depending on its meaning.

2.6 An initial cluster with ร, ถ, or ง presents problems in reading and pronunciation whether ร, ถ, or ง is a cluster with the preceding character or is a non-cluster, for example: เปรี่ยญ pârian and เปรี่ญ prâap.

2.7 A number of Thai words contain a silent ร or ิ, such as กิม sôt, เกียรติ kiat, จริง ciŋ, ใคร sây, หลวง suâng, ศิระ sâisâ, พรม phrom, พระมนุษย์ phraam.

2.8 A certain number of word finals contain either a silent vowel โ or โอ, such as ขัต châat, ประวัต prâwât, สมมต sômmút, ตาท thâat, เนื้อ hêet, but the two vowels are not pronounced.

2.9 Several words in Thai contain โ in the middle of the word, such as รห sâan, โหม โร om.

3. The existing systems of syllabification

3.1 The use of โ (yamakkan) over the first member of a consonant cluster and a final. This practice is widely adopted in Buddhist texts and it is still used in the Book of Prayer for Thai Buddhists. When a consonant-combination appears in Pali texts, a
yamkkan is placed on top of the first member of the combination, for example: ตั้วการะยณะ ฯ ด้วดอสะยนā.

A yamkkan or a silent marker, ęż, is also used in non-religious texts to signify a final. This system can be found in the manuscript of King Borommakot’s version of Chindamani where we doubt its genuineness and suspect that the manuscript was created in the early Bangkok period by a distinguished scholar with an intention to honor King Borommakot or to use his name as a pretext, for example: คณ khon, คณ khan, คณ khaan, คณ khin, คณ khiin, คณ khin, คณ khiin, คณ khun, คณ khuun, เคณ kheen, แคน khāēn, ไคน khoon, ครณ khoon, คณ khou, คณ khiian, เคณ khian, ไคณ khiian, เคณ khoon, ครณ khoon, คณ khou, คณ khoon.

3.2 The use of a dot below, to replace ęż, a yamkkan. It can be easily seen that the yamkkan takes up space above a character and is troublesome in writing. The use of a dot below seems to be neater in a certain aspect. This system is widely adopted in writing the Pali texts with Thai scripts up to present day, for example:

<table>
<thead>
<tr>
<th>สมมุน พุทฒิณ เขา</th>
<th>ราชวิศิษฎากร</th>
</tr>
</thead>
<tbody>
<tr>
<td>ดมณุษ นากะ จั่ป</td>
<td>สุขอนันทา ยกรำง</td>
</tr>
<tr>
<td>samane brāhmaṇe ceva</td>
<td>rājavamsikanātayo</td>
</tr>
<tr>
<td>amacce nāgare cápi</td>
<td>sanganhāti yathāraham</td>
</tr>
</tbody>
</table>

3.3 Both systems are adopted into the Thai writing system, but they do not completely solve syllabification problems of Thai for romanization and data processing. Problem 2.4 above can be easily solved, but

3.3.1 It is impossible to predict whether บูร is to be pronounced buan or būwcoon (see 2.1 and 2.5).

3.3.2 It is impossible to tell whether a character with a yamkkan on top, or a dot below is a final or the first member of a cluster (see 2.3 above). พอฤย can be read as phūō klaay or phōklaay depending on its meaning.

3.3.3 If we employ one of the existing systems, it is impossible to tell whether a character with a yamkkan above or a dot below in the following words is a final or

---

1 It is believed that the lower part of a yamkkan covers the first member of a cluster and the top part of it covers the second member of a cluster.
the first member of a cluster, such as เกษตรศาสตร์ kasèttrasat, มัตรฐาน mattrathaan, and นิทรา nitthraa.

4. The proposed systems for Thai spelling reform for data processing

4.1 The use of ´ to signify a cluster and , to signify a final. If a yamakkan is placed on top of the first member of a cluster (in order to avoid interfering with a tone mark or a vowel written above the line such as ˆ, ˇ, ˙, ˚) and a dot below is placed under the final, Thai words will be more facilitated for data processing and pronunciation. This system will face difficulty in placing a dot below a consonant when a final already has a vowel , written under the line.

4.1.1 The problems in 2.1 will be somewhat solved but not completely. Extra pronunciation rules must be devised such as:

4.1.1.1 If น, ภ, ผ, ฝ, น, ม, ฌ, ฉ, ด, or ส stands next to a single ธ which is followed by another consonant which is not a final, คะ is inserted after it; and if it stands next to a single ธ which is followed by a vowel, า is inserted after it, for example: กระท่ำ kóorákòt, กระท่ำก้าด้ากขอน, กระท่ำ khòorânni, กระท่ำ khârâwâat, กระท่ำ còorákaa, กระท่ำ câraacoon, กระท่ำ thoòorânni, กระท่ำ thâraathoon, กระท่ำ kòorâchëet, กระท่ำ náraathip, กระท่ำ mcoorâddèk, กระท่ำ mâru, กระท่ำ prâorâcàk, กระท่ำ wârum, สงสิทธิ์ สัตว์ râsît, ฟั่น sàriirâ, ฟั่น hòorânni, ฟั่น hârin, ฟั่น ?corânnût, ฟั่น ?ârin.

4.1.1.2 If ม precedes any consonant, except ธ, which is followed by a vowel or a covert vowel, คะ is inserted after it, for example: บรม boorrom, บริษัท booribuun, บริษัท boodin, บริษัท booribuan.

4.1.2 Problem 2.2 and 2.9 will be solved if character deletion rules are provided (see 2.2 and 3.3.3). It must be noted that the character deletion rules must come earlier in rule ordering.

4.1.3 Problems 2.3, 2.4, 2.5, and 2.6 can be completely solved and 2.7 can be partly solved.

---

2It must be noted that บวร is pronounced boorrom, but บริษัท is pronounced booribuan. Even บวร appears in the latter part of บริษัท, they are pronounced differently.