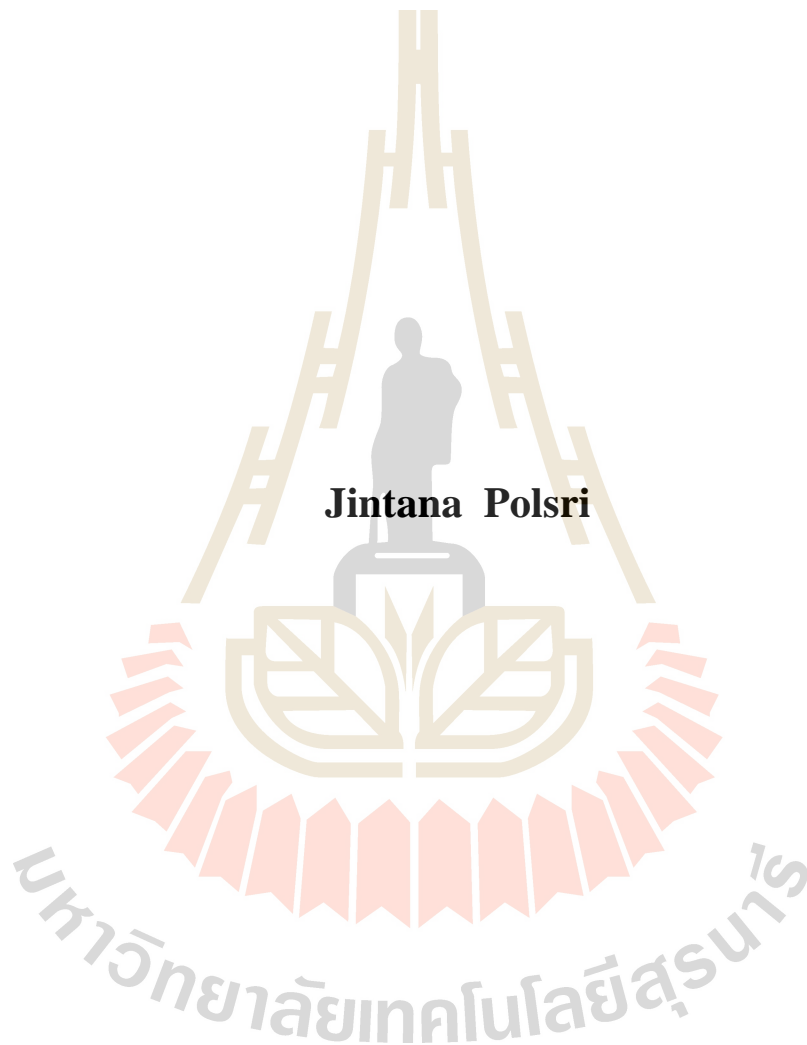


การพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อย
เป็นคำอ่านภาษาไทยถิ่นอีสาน



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2564

**THE DEVELOPMENT OF A FRAMEWORK FOR
TRANSLITERATING THAI NOI CHARACTERS INTO
ISAN LANGUAGE PRONUNCIATION**



**A Thesis Submitted in Partial Fulfilment of the Requirements for the
Degree of Doctor of Philosophy Program in Information Technology**


Suranaree University of Technology

Academic Year 2021


การพัฒนากรอบการทำงานในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่าน
ภาษาไทยถิ่นอีสาน

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้บัณฑิตวิทยาลัยรับผิดชอบนี้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปรัชญาดุษฎีบัณฑิต


คณะกรรมการสอบวิทยานิพนธ์




(รองศาสตราจารย์ ดร. จิตมนต์ อังสกุล)
ประธานกรรมการ



(ผู้ช่วยศาสตราจารย์ ดร. นิสาชล จำนงศรี)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)




(รองศาสตราจารย์ ดร. นิตยา เกิดประสพ)
กรรมการ




(ผู้ช่วยศาสตราจารย์ ดร. สุกกฤษฎี นิวัฒนา)
กรรมการ




(ดร. ชูชาติ หุชชะศักดิ์)
กรรมการ



(รองศาสตราจารย์ ดร. จีรยุทธ ไชยจารุวัฒน์)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม)



(รองศาสตราจารย์ ดร. นัตถชัย โชติชัยยางกูร)
รองอธิการบดีฝ่ายวิชาการและประกันคุณภาพ



(รองศาสตราจารย์ ดร. ธรา อังสกุล)
คณบดีสำนักวิทยาศาสตร์และศิลปประดิษฐ์

จินตนา พลศรี : การพัฒนากรอบการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อย เป็นคำอ่านภาษาไทยถิ่นอีสาน (THE DEVELOPMENT OF A FRAMEWORK FOR TRANSLITERATING THAI NOI CHARACTERS INTO ISAN LANGUAGE PRONUNCIATION) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.นิสาชล จำนงศรี, รองศาสตราจารย์ ดร.จิรยุทธ ไชยจรรวมิช, 218 หน้า.

การวิจัยนี้ นำเสนอกรอบการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน วิจัยดำเนินการวิจัยแบ่งออกเป็น 3 ระยะ ได้แก่ ระยะที่ 1 ประกอบด้วยการศึกษาปัญหาในการปรัวรรตเอกสาร โบราณ การศึกษาโครงสร้างการบันทึกคำรายอีสาน โบราณที่บันทึกในใบลานด้วยอักษรไทยน้อยที่ส่งผลต่อประสิทธิภาพการปรัวรรต และลักษณะอักษร ภาษา และอักษรวิธอักษรไทยน้อยที่ใช้บันทึกคำรายอีสาน โบราณ วิจัยใช้วิธีวิจัยเชิงคุณภาพ รวบรวมข้อมูลโดยการสัมภาษณ์นักภาษาศาสตร์โบราณ จำนวน 10 คน ด้วยแบบสัมภาษณ์แบบกึ่งโครงสร้าง และการวิเคราะห์คำรายอีสานที่ปรัวรรตเอาไว้แล้วพร้อมภาพคำรายต้นฉบับ จำนวน 11 ผูก 459 หน้าลาน ผลการศึกษาพบว่า ปัญหาการปรัวรรตเอกสาร โบราณและคำรายอีสานมีลักษณะคล้ายคลึงกัน และที่เป็นปัญหาอยู่ในระดับมากมีสาเหตุจากลักษณะทางกายภาพของเอกสารโบราณ รูปอักษร ภาษา และอักษรวิธของอักษรไทยโบราณ การบันทึกคำรายอีสานจะบันทึกด้วย 2 ภาษาปะปนกันระหว่างภาษาไทยถิ่นอีสาน และบาลี อักษรไทยน้อยมีจำนวนตัวอักษรน้อยกว่าอักษรไทยปัจจุบันและไม่มีวรรณยุกต์ การประสมอักษรเป็นคำส่วนมากคล้ายกับอักษรไทยปัจจุบัน และบางส่วนคล้ายกับอักษรธรรมอีสาน คำเดียวกันสะกดได้หลายแบบ และความสัมพันธ์ระหว่างคำปรัวรรตของอักษรไทยน้อยและคำอ่านภาษาไทยถิ่นอีสานมี 4 รูปแบบ คือ 1:1 1:M M:1 M:N ทำให้การปรัวรรตอักษรไทยน้อยต้องอาศัยบริบทและเข้าใจภาษาไทยถิ่นอีสาน ระยะที่ 2 การออกแบบและพัฒนากรอบการทำงานในการแปลงคำปรัวรรตของตัวอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ผลการวิจัยประกอบด้วย ได้แก่ (1) คลังศัพท์คำปรัวรรต ขนาด 4,645 คำ ใช้สำหรับตัดคำ และค้นหาคำอ่านภาษาไทยถิ่นอีสานที่เป็นไปได้ทั้งหมด และ (2) ออนโทโลยีคำศัพท์ใช้สำหรับการค้นหาคำอ่านภาษาไทยถิ่นอีสานที่มีความกำกวมและต้องใช้คำบริบทในการเจาะจงคำอ่านที่เหมาะสม โครงสร้างออนโทโลยีประกอบด้วย 20 คลาสหลัก และ (3) กระบวนการและขั้นตอนวิธีแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน แบ่งเป็น 3 ขั้นตอน ได้แก่ การตัดคำ การค้นหาคำอ่านด้วยคลังศัพท์ และการค้นหาคำอ่านด้วยออนโทโลยี และระยะที่ 3 การประเมินประสิทธิภาพของกรอบการทำงาน ผลการประเมินในเชิงปริมาณมีความถูกต้องในการแปลงคำปรัวรรตของตำรับยาโบราณอีสาน จำนวน 807 ตำรับ (7,929 คำ) เป็นคำอ่าน

ภาษาไทยถิ่นอีสาน มีความถูกต้องร้อยละ 87.84 และผลการประเมินประสิทธิภาพเชิงคุณภาพโดยนักภาษาโบราณ จำนวน 12 คน และผู้ใช้ประโยชน์จากตำรายาโบราณอีสาน 27 คน มีค่าเฉลี่ยทุกรายการอยู่ในระดับดี ($\bar{X} = 4.08$, $SD = 0.73$)



สาขาวิชาเทคโนโลยีสารสนเทศ

ปีการศึกษา 2564

ลายมือชื่อนักศึกษา Am ant.

ลายมือชื่ออาจารย์ที่ปรึกษา อ.อ.อ.

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม อ.อ.อ.



JINTANA POLSRI : THE DEVELOPMENT OF A FRAMEWORK FOR
TRANSLITERATING THAI NOI CHARACTERS INTO ISAN LANGUAGE
PRONUNCIATION. THESIS ADVISOR : ASST. PROF. NISACHOL
CHAMNONGSRI, Ph.D., ASSOC. PROF. JEERAYUT CHAJARUWANICH,
Ph.D., 218 PP.

ONTOLOGY/ TRANSLITERATING/ THAI NOI CHARACTER/ ISAN LANGUAGE

This research presented the framework to transliterate Thai Noi characters into Isan language pronunciations. This research was conducted into 3 phases summarized as follows: Phase 1 consisted of the studies of the problems of ancient document transliteration, the recording structures of ancient Isan medicine textbooks inscribed on palm leaves with Thai Noi characters affecting the efficiency of transliterating, and the character characteristics, language, and Thai Noi characters inscribed on the ancient Isan medicine textbooks. The research methodology used was the qualitative research method. Data were collected from interviewing 10 ancient language linguists using semi-structured interview forms, analyzing the ancient Isan medicine textbooks already transliterated, and 11 copies of the original digital pictures with 459 pages. The results showed that the problems of transliterating ancient documents and ancient Isan medicine textbooks were similar. The problems found were at a high level and from the physical characteristics of the ancient documents, the orthography of the ancient Thai scripts having no tonal markers, In terms of the recording structures and characteristics of Thai Noi characters of ancient Isan medicine textbooks, 2 languages, Isan language and Pali, were combinedly used. The Thai Noi characters are fewer number than standard Thai characters. Their orthography was similar to that of the present Thai characters, but some were derived from the orthography of Dhamma Isan scripts.

A Thai Noi word could be spelled in several ways. There were no both any marks placed over the final consonant of the word and tonal markers. The relationship between the transliterating of Thai Noi characters and words of Isan pronunciation language could be divided into 4 forms: 1:1, 1: M, M:1, and M: N. affected the context and understanding of Isan language was very important for transliterating Thai Noi words into Isan pronunciation language. Phase 2 was the design and development of a framework of Thai Noi characters into Isan pronunciation language. The components of the framework were described as follows: (1) The lexicon of 4,645 ancient Isan medicine words consisted of transliterating words, Isan pronunciation language words and Thai standard words. The lexicon was used for text segmentation and to find all Isan pronunciation language words. (2) The lexicon ontology of ancient Isan medicine textbooks for query the ambiguous Isan pronunciation language words and require context to identify appropriate word consisted of 20 main classes and (3) The process and algorithm for converting the transliteration of Thai Noi characters into Isan pronunciation language words was divided into 3 steps: text segmentation, searching all possible Isan pronunciation language words by lexicon and searching Isan pronunciation language words by ontology. Phase 3 were the evaluation of the efficacy of the framework. The result showed that the accuracy of the transliteration of 807 ancient Isan medicines recipes (7,929 words) into the Isan pronunciation language was 87.8 percent and the evaluate the framework efficacy have overall mean was good level in all items ($\bar{x} = 4.08, SD = 0.73$).

School of Information Technology
Academic Year 2021

Student's Signature 
 Advisor's Signature 
 Co-Advisor's Signature 

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้เนื่องด้วยความกรุณาและการดูแลเอาใจใส่ของผู้ช่วยศาสตราจารย์ ดร.นิศาชล จำนงศรี อาจารย์ที่ปรึกษาวิทยานิพนธ์ และรองศาสตราจารย์ ดร.จิรยุทธ ไชยจรรวม อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่คอยให้คำแนะนำ คำปรึกษา แนะนำแนวทางในการทำวิจัย ตลอดจนดูแลเวลาอันมีค่าในการตรวจทานวิทยานิพนธ์ให้มีความถูกต้องสมบูรณ์ ผู้วิจัยขอขอบพระคุณอย่างสูงมา ณ โอกาสนี้

ขอขอบพระคุณ รองศาสตราจารย์ ดร.จิตติมนต์ อังสกุล ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. นิตยา เกิดประสพ ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษณ์ นิวัฒนากุล ดร.ชูชาติ หลุยชัยศักดิ์ กรรมการสอบวิทยานิพนธ์ ที่ได้เสียสละเวลาอันมีค่าในการพิจารณาและให้คำแนะนำในการปรับปรุงแก้ไขที่เป็นประโยชน์ต่อวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณ ผู้เชี่ยวชาญ นักภาษาโบราณ นักวิชาการ จาก กลุ่มหนังสือตัวเขียนและจารึก หอสมุดแห่งชาติ ท้าวสุกรี กรุงเทพมหานคร และสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม ทุกท่าน ที่ให้ความกรุณาเสียสละเวลาอันมีค่าในการให้ความรู้และคำปรึกษา ด้านการปริวรรตเอกสาร โบราณและตำรายาอีสานโบราณ ตลอดจนอำนวยความสะดวกในการเข้าไปศึกษาค้นคว้า นับตั้งแต่วันแรกจนกระทั่งงานวิจัยสำเร็จเป็นเวลายาวนานกว่า 4 ปี จนกระทั่งงานวิจัยแล้วเสร็จ วิทยานิพนธ์นี้จะสำเร็จไม่ได้เลยหากไม่ได้รับความกรุณาเหล่านี้ โดยเฉพาะสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน ที่อนุญาตให้นำเอาตำรายาอีสานโบราณของโครงการมาใช้ในงานวิจัย ตรวจทานความถูกต้องของข้อมูลที่ใช้ในงานวิจัย รวมถึง คณะแพทย์ศาสตร์ มหาวิทยาลัยมหาสารคาม ที่ให้ความอนุเคราะห์และอนุญาตให้นำเอาผลการวิจัยไปให้นักศึกษาสาขาวิชาการแพทย์แผนไทยประยุกต์ช่วยทำการประเมินประสิทธิภาพงานวิจัย ทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงตามวัตถุประสงค์

ท้ายนี้ ขอกราบขอบพระคุณบิดาและมารดาที่เป็นผู้วางรากฐานชีวิต ให้โอกาสทางการศึกษาและเป็นกำลังใจให้ข้าพเจ้าในทุก ๆ ด้าน ขอขอบคุณ ครอบครัวเพื่อน พี่น้อง กัลยาณมิตร ทุกท่าน ที่ให้การสนับสนุน ช่วยเหลือ เป็นกำลังใจ อย่างสม่ำเสมอในทุกโอกาส รวมถึงมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ หน่วยงานต้นสังกัด ที่ให้การสนับสนุนด้านทุนการศึกษาในครั้งนี้

จินตนา พลศรี

สารบัญ

	หน้า
บทคัดย่อ (ภาษาไทย).....	ก
บทคัดย่อ (ภาษาอังกฤษ).....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฎ
สารบัญรูป.....	ฅ
บทที่	
1 บทนำ	1
1.1 ความเป็นและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์การวิจัย.....	6
1.3 คำถามการวิจัย.....	6
1.4 ขอบเขตการวิจัย.....	6
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	7
1.6 คำอธิบายศัพท์.....	8
2 ปรัชมนวัตกรรมและงานวิจัยที่เกี่ยวข้อง	9
2.1 อักษรไทยน้อย.....	10
2.1.1 ความเป็นมาของอักษรไทยน้อย.....	10
2.1.2 อักษรวิธีอักษรไทยน้อย.....	11
2.2 การปริวรรตอักษรไทยน้อย.....	15
2.2.1 การปริวรรตอักษรโบราณ.....	15
2.2.2 หลักเกณฑ์การปริวรรตเอกสารโบราณ.....	15
2.2.3 หลักเกณฑ์และขั้นตอนการปริวรรตอักษรไทยน้อย.....	17
2.3 ภาษาไทยถิ่นอีสาน.....	19
2.3.1 ลักษณะภาษาไทยถิ่น.....	19
2.3.2 ไวยากรณ์ภาษาไทยถิ่นอีสาน.....	20
2.3.3 การประสมคำภาษาไทยถิ่นอีสาน.....	29

สารบัญ (ต่อ)

	หน้า
2.3.4 ประโยคภาษาไทยถิ่นอีสาน.....	33
2.4 ออนโทโลยี.....	34
2.4.1 ความหมายและความสำคัญของออนโทโลยี.....	34
2.4.2 องค์ประกอบของออนโทโลยี.....	35
2.4.3 ประเภทของออนโทโลยี.....	36
2.4.4 กระบวนการพัฒนาออนโทโลยี.....	37
2.4.5 ภาษาที่ใช้ในการพัฒนาออนโทโลยี.....	40
2.4.6 โปรแกรมสร้างออนโทโลยี.....	42
2.4.7 การประเมินออนโทโลยี.....	43
2.5 คลังศัพท์.....	45
2.5.1 ความหมาย.....	45
2.5.2 การรวบรวมคำศัพท์เพื่อสร้างคลังศัพท์.....	46
2.5.3 ขั้นตอนการสร้างคลังศัพท์.....	47
2.5.4 คลังศัพท์อิเล็กทรอนิกส์.....	48
2.6 การประมวลผลภาษาธรรมชาติ.....	50
2.6.1 ความหมายของการประมวลผลภาษาธรรมชาติ.....	50
2.6.2 กระบวนการทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติ.....	50
2.6.3 ขั้นตอนการประมวลผลภาษาธรรมชาติ.....	51
2.7 การแปลภาษาด้วยเครื่อง.....	53
2.7.1 ความหมายของการแปลภาษาด้วยเครื่อง.....	53
2.7.2 ประเภทของระบบแปลภาษาด้วยเครื่อง.....	53
2.8 วรรณกรรมที่เกี่ยวข้อง.....	55
2.8.1 การแปลภาษาไทยโบราณด้วยเครื่อง.....	55
2.8.2 การพัฒนาคลังศัพท์อิเล็กทรอนิกส์สำหรับงานประมวลผลภาษาธรรมชาติ.....	58
2.8.3 การใช้ออนโทโลยีในการประมวลผลภาษาธรรมชาติ.....	59
2.9 กรอบแนวคิดในการวิจัย.....	64

สารบัญ (ต่อ)

	หน้า
3	วิธีดำเนินการวิจัย 66
3.1	ระยะที่ 1 การศึกษาปัญหาในการปริวรรตตำรายอีสาน โบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย..... 66
3.1.1	การศึกษากระบวนการปริวรรตเอกสาร โบราณและปัญหาในการปริวรรตตำรายอีสาน โบราณที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย..... 66
3.1.2	การศึกษาโครงสร้างการบันทึกตำรายอีสาน โบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย ที่ส่งผลกระทบต่อประสิทธิภาพในการปริวรรต..... 67
3.1.3	การศึกษาลักษณะของอักษรไทยน้อย และอักษรวิธีของอักษรไทยน้อยที่ใช้บันทึกตำรายอีสาน โบราณในใบลาน..... 68
3.2	ระยะที่ 2 การออกแบบและพัฒนากรอบแนวคิดในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 69
3.2.1	การออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน..... 69
3.2.2	การพัฒนาออนไลน์คำอ่านตำรายอีสาน โบราณ..... 73
3.2.3	การออกแบบขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 74
3.2.4	การสร้างข้อมูลชุดทดสอบประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 74
3.3	ระยะที่ 3 การประเมินประสิทธิภาพกรอบการทำงานทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 75
3.3.1	การประเมินประสิทธิผลของกรอบแนวคิดในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 75
3.3.2	การประเมินประสิทธิภาพกรอบแนวคิดในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน..... 75
4	ผลการศึกษาปัญหาการปริวรรตตำรายอีสานโบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย 79
4.1	การศึกษากระบวนการปริวรรตเอกสาร โบราณ และปัญหาในการปริวรรตตำรายอีสานโบราณที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย..... 79

4.1.1	ข้อมูลส่วนตัว ประสบการณ์ และความเชี่ยวชาญด้านการปริวรรต เอกสารโบราณ.....	80
4.1.2	ปัญหาการปริวรรตเอกสารโบราณ.....	82
4.1.3	กระบวนการปริวรรตเอกสารโบราณ.....	84
4.1.4	ปัญหาการปริวรรตตำรายอีสานโบราณที่บันทึกในใบลานด้วยอักษร ไทยน้อย.....	86
4.2	การศึกษาโครงสร้างการเขียนตำรายอีสานโบราณที่บันทึกในใบลานด้วย อักษรไทยน้อยที่ส่งผลกระทบต่อประสิทธิภาพในการปริวรรต.....	89
4.2.1	องค์ประกอบด้านเนื้อหาที่บันทึกในตำรับยาโบราณอีสาน.....	89
4.2.2	โครงสร้างการเขียนตำรายาโบราณอีสาน.....	92
4.2.3	การจัดกลุ่มคำศัพท์ที่ใช้เขียนในตำรายอีสานโบราณ.....	101
4.3	การศึกษาลักษณะอักษรไทยน้อย และอักษรวิธอักษรไทยน้อยที่ใช้บันทึกตำรายา โบราณอีสานในใบลาน.....	106
4.3.1	รูปอักษรไทยน้อย.....	106
4.3.2	ภาษา.....	111
4.3.3	อักษรวิธีการประสมอักษรไทยน้อยเป็นคำ.....	112
4.3.4	ลักษณะการสะกดคำของอักษรไทยน้อย.....	113
4.4	ลักษณะอักษรไทยน้อยที่ส่งผลกระทบต่อปริวรรตตำรายอีสานโบราณแบบ อัติโนมติ.....	118
4.4.1	ปัญหาการทำคำปริวรรต.....	118
4.4.2	ปัญหาการทำคำอ่านภาษาไทยถิ่นอีสาน.....	118
5	ผลการพัฒนาและประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของ อักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	121
5.1	การพัฒนาคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน.....	122
5.1.1	ลักษณะคลังคำภาษาไทยถิ่นอีสานสำหรับการแปลงคำปริวรรตเป็นคำอ่าน ภาษาไทยถิ่นอีสานแบบอัติโนมติ.....	122
5.1.2	การวิเคราะห์และออกแบบโครงสร้างคลังคำ คำปริวรรต-คำอ่านภาษาไทยถิ่น อีสาน-คำอ่านภาษาไทยมาตรฐาน.....	124
5.1.3	เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสาน.....	129
5.1.4	การพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทย มาตรฐาน.....	138

สารบัญ (ต่อ)

	หน้า
5.2 การพัฒนาออนไลน์คำราชาอีสานโบราณ.....	149
5.2.1 ความต้องการออนไลน์สำหรับแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	149
5.2.2 พัฒนาออนไลน์คำราชาอีสานโบราณ.....	150
5.2.3 ผลการประเมินประสิทธิภาพของออนไลน์คำราชาอีสานโบราณ.....	155
5.2.4 ผลการประเมินประสิทธิภาพออนไลน์คำศัพท์คำราชาอีสานโบราณ.....	156
5.3 การออกแบบขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	158
5.3.1 การตัดคำ.....	160
5.3.2 การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์.....	161
5.3.3 การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยออนไลน์คำราชาอีสานโบราณ.....	163
5.4 การประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	165
5.4.1 ความถูกต้องแม่นยำ.....	165
5.4.2 ความผิดพลาด.....	166
5.4.3 ประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	167
6 สรุปอภิปรายผล และข้อเสนอแนะ.....	175
6.1 สรุปผลการวิจัย.....	175
6.1.1 ระยะที่ 1 การศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตคำราชาอีสานโบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย.....	175
6.1.2 ระยะที่ 2 การออกแบบและพัฒนากรอบแนวคิดในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	180
6.1.3 ระยะที่ 3 การประเมินประสิทธิผลของกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	185
6.2 อภิปรายผลการวิจัย.....	187
6.2.1 กรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	187

6.2.2	คลังศัพท์.....	188
6.2.3	ออนโทโลยีคำศัพท์.....	191
6.3	ข้อจำกัดการวิจัย.....	192
6.4	การประยุกต์ผลการวิจัย.....	192
6.5	ข้อเสนอแนะในการวิจัยครั้งต่อไป.....	193
6.5.1	การเพิ่มประสิทธิภาพคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน- คำอ่านภาษาไทยมาตรฐาน.....	193
6.5.2	ออนโทโลยีคำราชาโบราณอีสาน.....	193
6.5.3	คลังข้อความคู่ขนาน คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทย มาตรฐาน.....	194
6.5.4	ระบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาอีสาน.....	194
6.6	องค์ความรู้ใหม่ในงานวิจัย (Contribution).....	194
	รายการอ้างอิง.....	196
	ภาคผนวก	
ภาคผนวก ก	แบบฟอร์มบันทึกการสัมภาษณ์.....	203
ภาคผนวก ข	แบบประเมินประสิทธิภาพการทำงานในการแปลงคำปริวรรต ของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	215
	ประวัติผู้เขียน.....	218

สารบัญตาราง

ตารางที่	หน้า
2.1	งานวิจัยด้านการปริวรรตเอกสาร โบราณของไทยด้วยเครื่อง..... 61
4.1	หมวดเนื้อหาในเอกสาร โบราณที่นักภาษาโบราณเคยปริวรรต..... 81
4.2	จำนวนนักภาษาโบราณและระดับความเชี่ยวชาญการปริวรรตเอกสาร โบราณ จำแนกตามอักษร โบราณที่อ่านได้..... 81
4.3	ความเชี่ยวชาญด้านการอ่านภาษาโบราณของนักภาษาโบราณ..... 81
4.4	ระดับปัญหาด้านภาษาและอักษรวิธีอักษรไทย โบราณที่ส่งผลต่อการปริวรรต อักษรโบราณ..... 83
4.5	รายละเอียดคำราชาอีสาน โบราณที่บันทึกในใบลานด้วยอักษรไทยน้อยที่นำมา ศึกษา..... 89
4.6	รูปพยัญชนะตัวสะกดอักษรไทยน้อยและตัวอย่างคำศัพท์แบ่งตามมาตราตัวสะกด..... 107
4.7	รูปพยัญชนะประสมของอักษรไทยน้อยและตัวอย่างคำศัพท์..... 108
4.8	รูปพยัญชนะควบกล้ำในอักษรไทยน้อยและตัวอย่างคำศัพท์..... 108
4.9	ตัวอย่างรูปสระอักษรไทยน้อยและตัวอย่างการใช้..... 109
4.10	ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่พ้องเสียงกับคำศัพท์ภาษาไทยมาตรฐาน..... 112
4.11	ตัวอย่างคำศัพท์ที่สะกดด้วยอักษรไทยน้อยที่ไม่ใช่พยัญชนะควบกล้ำและตัวการันต์..... 113
4.12	ตัวอย่างการสะกดคำด้วยอักษรไทยน้อย กรณีคำเดียวกันใช้ทั้งสระเสียงสั้นและสระ เสียงยาวในการสะกด..... 114
4.13	ตัวอย่างการสะกดคำด้วยอักษรไทยน้อยผสมกับอักษรธรรมอีสาน..... 114
4.14	ตัวอย่างการเขียนตัวเลขของอักษรไทยน้อยในคำราชาอีสาน โบราณ..... 116
4.15	ตัวอย่างการสะกดคำด้วยอักษรวิธีพิเศษที่ขึ้นอยู่กับผู้เขียนแต่ละคน..... 116
4.16	ตัวอย่างการวางสระไม่ตรงตำแหน่งพยัญชนะต้นที่เป็นปัญหาในการอ่าน..... 117
5.1	ผลการวิเคราะห์โครงสร้างการอธิบายคำศัพท์ของคลังคำภาษาไทยมาตรฐานและ ภาษาไทยถิ่นอีสาน..... 123
5.2	พจนานุกรมข้อมูลเอ็นทีคำศัพท์ (Vocabulary Data Dictionary)..... 126
5.3	พจนานุกรมข้อมูลเอ็นทีคำปริวรรต (Transliteration Dictionary)..... 126

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
5.4 พจนานุกรมข้อมูลเอ็นทีดีแหล่งข้อมูลอ้างอิง (Referenced resources Dictionary)	126
5.5 พจนานุกรมข้อมูลเอ็นทีดีคำปริวรรต-คำอ่าน (TransHas Vocab Dictionary).....	126
5.6 ชนิดคำในคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน.....	127
5.7 ตัวอย่างคำภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงหรือคล้ายกับภาษาไทยมาตรฐาน และเป็นคำที่มีความหมายเดียวกัน.....	130
5.8 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกัน หรือคล้ายกับภาษาไทย ที่มาตรฐานเป็นคำที่มีตัวสะกดการันต์ และมีความหมายเดียวกัน.....	130
5.9 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกันหรือคล้ายกับภาษาไทย มาตรฐานที่เป็นคำมีพยัญชนะควบกล้ำ.....	131
5.10 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงเป็นปฏิภาคกับหน่วยเสียง “จ” “ช” “ฉ” หรือ “ส” ในภาษาไทยมาตรฐาน.....	132
5.11 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “ฮ” ปฏิภาคกับหน่วยเสียง “ร” ในภาษาไทยมาตรฐาน.....	132
5.12 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง ข-นาสิก.....	133
5.13 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “อย”.....	134
5.14 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานและภาษาไทยถิ่นอีสานที่มีความหมายเดียวกัน แต่มีหน่วยเสียงสระหรือพยัญชนะต่างกันเล็กน้อย.....	134
5.15 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานออกเสียงต่างจากภาษาไทยมาตรฐาน.....	135
5.16 เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังคำปริวรรต-คำอ่านภาษาไทย ถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน.....	136
5.17 เกณฑ์การแปลงคำอ่านภาษาไทยถิ่นอีสานเป็นคำปริวรรต.....	141
5.18 สรุปจำนวนคำศัพท์จากตำรายาอีสานโบราณที่ปริวรรตจากอักษรไทยน้อย.....	143
5.19 จำนวนคำศัพท์ภาษาไทยถิ่นอีสานจำแนกตามประเภทความสัมพันธ์.....	144
5.20 รายชื่อผู้เชี่ยวชาญการปริวรรตตำรายาอีสานโบราณอักษรไทยน้อย.....	146
5.21 ผลการประเมินประสิทธิภาพคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่าน ภาษาไทยมาตรฐาน.....	155
5.22 Data Properties.....	155

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
5.23	Object Properties.....	158
5.24	ผลการประเมินประสิทธิภาพออนไลน์โดยสำรวจอีสานโบราณ.....	166
5.25	ผลการประเมินความถูกต้องกรอบการทำงานในการแปลงคำปริวรรตของอักษร ไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	173
5.26	ข้อมูลพื้นฐานของผู้ประเมิน.....	174
5.27	ผลการประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษร ไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	185



สารบัญรูปลูกภาพ

รูปที่	หน้า
2.1	พยัญชนะเดี่ยวอักษรไทยน้อย..... 12
2.2	พยัญชนะประสมและควบกล้ำอักษรไทยน้อย..... 12
2.3	รูปสระอักษรไทยน้อย..... 13
2.4	รูปตัวเลขอักษรไทยน้อย..... 13
2.5	ตัวอย่างวิธีการประสมอักษรไทยน้อยเป็นคำ..... 14
2.6	ตัวอย่างการปรัวรรณอักษรไทยน้อย..... 18
2.7	กรอบแนวคิดการวิจัย..... 65
4.1	กระบวนการปรัวรรณเอกสารโบราณ และผลลัพธ์การปรัวรรณในแต่ละขั้นตอน..... 85
4.2	ลักษณะกายภาพและการบันทึกตำรายาอีสานโบราณในเอกสารโบราณ..... 86
4.3	ลักษณะการบันทึกตำรายาอีสานโบราณในโบราณ..... 87
4.4	แผนภาพทักษะที่ต้องใช้ในการปรัวรรณตำรายาอีสานโบราณที่บันทึกด้วยอักษรไทยน้อยในโบราณ..... 88
4.5	ตัวอย่างองค์ประกอบด้านเนื้อหาในตำรายาอีสานโบราณ..... 91
4.6	ตัวอย่างองค์ประกอบด้านเนื้อหาในตำรายาอีสานโบราณ..... 92
4.7	ตัวอย่างโครงสร้างตำรับยาประเภทใช้สมุนไพรที่รักษาได้ 1 โรค มี 1 ขนาน..... 93
4.8	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค มีขนานหลัก และขนานทางเลือก..... 94
4.9	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค มีขนานหลัก และขนานทางเลือกสำหรับใช้เมื่อขนานหลักใช้ไม่ได้ผล..... 94
4.10	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค และขนานย่อยใช้ตามลำดับ..... 95
4.11	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรคได้หลายโรคด้วยขนานเดียวกัน..... 95
4.12	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่มีหลายขนาน แต่ละขนานใช้ตามอาการหรือใช้เมื่อขนานก่อนหน้าใช้ไม่ได้ผล..... 96
4.13	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่มีคำรักษา (กาย)..... 96
4.14	ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่บอกอาการของโรคและชื่อโรค..... 97

สารบัญรูปร่างภาพ (ต่อ)

รูปที่	หน้า
4.15 ตัวอย่างโครงสร้างคำรับยาสมุนไพรที่บอกชื่อโรค อาการ และชื่อเรียกอาการ.....	97
4.16 ตัวอย่างโครงสร้างคำรับยาสมุนไพรที่รักษา 1 โรค มีหลายขนานให้ใช้ตามเดือนที่ป่วย.....	98
4.17 ตัวอย่างโครงสร้างคำรับยาที่ใช้สมุนไพรร่วมกับพิธีกรรมหรือคาถาในการรักษา.....	99
4.18 ตัวอย่างโครงสร้างคำรับยาประเภทรักษาโรคด้วยคาถา.....	100
4.19 ตัวอย่างโครงสร้างเนื้อหาประเภทอื่น ๆ ในตำรายาอีสานโบราณ.....	100
4.20 อนุกรมวิธานคำศัพท์ในตำรายาอีสานโบราณ.....	105
4.21 ตัวอย่างรูปพยัญชนะอักษรไทยน้อยที่ใช้เป็นพยัญชนะต้น.....	106
4.22 ตัวอย่างรูปตัวเลขในอักษรไทยน้อยในตำรายาอีสานโบราณ.....	108
4.23 เครื่องหมายจบคำรับยาอีสานโบราณ.....	110
4.24 ตัวอย่างคาถาภาษาบาลี-สันสกฤตสำหรับเสกใส่ยา.....	111
4.25 การประสมอักษรไทยน้อยเป็นคำ.....	112
4.26 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อย กรณีคำเดียวกันแต่ใช้พยัญชนะต้นต่างกัน.....	114
4.27 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อยที่ไม่มีวรรณยุกต์.....	115
4.28 ตัวอย่างการแก้ไขตัวสะกดในตำรายาอีสานโบราณ.....	117
4.29 ตัวอย่างคำอักษรไทยน้อยคำเดียวกันแต่สะกดคำต่างกัน.....	119
4.30 ตัวอย่างคำปริวรรตและคำอ่านภาษาไทยถิ่นอีสานจำแนกตามประเภทความสัมพันธ์.....	120
5.1 องค์ประกอบการอธิบายคำศัพท์ของคลังคำสำหรับตัดคำและแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน.....	125
5.2 แผนภาพเอ็นทีดีและความสัมพันธ์ของคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน.....	125
5.3 ขั้นตอนการสร้างฐานข้อมูลคลังคำศัพท์.....	138
5.4 โครงสร้างฐานข้อมูลคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (IsanMedDictionary) สร้างด้วยระบบจัดการฐานข้อมูล SQLite Studio.....	140
5.5 ตัวอย่างการตัดคำและกำกับชนิดของคำ.....	142
5.6 ตัวอย่างการลบคำซ้ำ.....	142
5.7 ตัวอย่างการแปลงคำอ่านภาษาไทยถิ่นอีสานเป็นคำปริวรรต.....	143

สารบัญรูปร่าง (ต่อ)

รูปที่	หน้า
5.8 ตัวอย่างการกำหนดรายละเอียดคำศัพท์ในคลังคำ	143
5.9 ขั้นตอนการนำข้อมูลคำศัพท์เข้าสู่ระบบฐานข้อมูล	145
5.10 ตัวอย่างข้อมูลคำศัพท์ที่ในฐานข้อมูลคลังคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน- คำอ่านภาษาไทยมาตรฐาน	145
5.11 ตัวอย่างการสอบถามคำอ่านภาษาไทยถิ่นอีสานและบริบทในออนโทโลยีตำรายา อีสานโบราณ	149
5.12 โครงสร้างออนโทโลยีตำรายาโบราณอีสาน	154
5.13 ตัวอย่างการกำหนดคุณลักษณะของคำศัพท์ (Properties Assertions)	154
5.14 กระบวนการค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยออนโทโลยีคำศัพท์ตำยาโบราณ อีสานและตัวอย่าง	156
5.15 ตัวอย่างการตรวจความถูกต้องของคำอ่านและการหาระดับความถูกต้องของคำอ่าน	157
5.16 กระบวนการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน	158
5.17 ขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน	159
5.18 ขั้นตอนการตัดคำปริวรรต	160
5.19 ตัวอย่างการตัดคำปริวรรตผิดพลาด	161
5.20 ตัวอย่างคำปริวรรตที่สามารถตัดคำได้หลายแบบ	161
5.21 ขั้นตอนการทำคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์	162
5.22 ตัวอย่างผลการค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคลังคำปริวรรต-คำอ่านภาษาไทย ถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน	162
5.23 ตัวอย่างคำอ่านภาษาไทยถิ่นอีสานที่มีคำอ่านได้หลายคำ และคำบริบทซ้าย-ขวา	163
5.24 ขั้นตอนวิธีการเลือกคำอ่านภาษาไทยถิ่นอีสานที่เหมาะสมด้วยออนโทโลยีคำศัพท์	164
5.25 ตัวอย่างการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน กรณี แปลงคำผิด	167
5.26 ตัวอย่างการตัดคำแบบคำประสมและการตัดคำแบบคำโดดที่ส่งผลต่อการแปลงเป็น คำอ่านภาษาไทยถิ่นอีสานและความหมายของคำ	167
5.27 ตัวอย่างหน้าจอหลักระบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทย ถิ่นอีสาน	168
5.28 ตัวอย่างหน้าจอปริวรรตตำรายาฉบับใหม่	169

สารบัญรูปภาพ (ต่อ)

รูปที่		หน้า
5.29	ตัวอย่างหน้าจอแสดงกรอกรายละเอียดข้อมูลตำรายาที่จะปริวรรตฉบับใหม่.....	170
5.30	ตัวอย่างหน้าจออัปโหลดภาพดิจิทัลโบราณตำรายาอีสาน โบราณที่จะปริวรรต.....	170
5.31	ตัวอย่างหน้าจอพิมพ์คำปริวรรต.....	171
5.32	ตัวอย่างหน้าจอแสดงรายชื่อตำรายาที่ปริวรรตและทำคำอ่านไว้แล้ว.....	171
5.33	ตัวอย่างหน้าจอรายละเอียดการแปลงคำปริวรรตเป็นคำอ่าน.....	172



บทที่ 1

บทนำ

1.1 ความเป็นมาและที่มาของปัญหาการวิจัย

เอกสารโบราณ หมายถึง เอกสารที่มีความเก่าแก่ อายุมากกว่า 100 ปีขึ้นไป (สำนักหอสมุดแห่งชาติ, 2551) มีหลายประเภท อาทิ จารึก หนังสือสมุดไทย เอกสารโบราณ คนโบราณใช้บันทึกเนื้อหาต่าง ๆ ทั้งทางโลกและทางธรรม เช่น หลักธรรมคำสอนทางพระพุทธศาสนา ประวัติศาสตร์ โหราศาสตร์ วรรณกรรม พงศาวดาร การปกครอง กฎหมายโบราณและจารีตประเพณี ตำรายา คธาอาคม เป็นต้น เนื้อหาเหล่านี้เป็นประโยชน์ต่อการนำไปศึกษาของคนในยุคปัจจุบันเป็นอย่างยิ่ง อย่างไรก็ตาม การนำเอกสารโบราณไปศึกษาค้นคว้าเป็นเรื่องที่ทำได้ยาก เนื่องจากเอกสารโบราณบันทึกด้วยอักษรและภาษาโบราณที่ปัจจุบันเลิกใช้แล้วจะต้องอาศัยผู้ที่มีความเชี่ยวชาญด้านอักษรและภาษาโบราณปริวรรตเป็นอักษรและภาษาไทยปัจจุบันเสียก่อน

การปริวรรต (Transliterations) เป็นกระบวนการถ่ายถอด แปลงอักษร ภาษา และอักษรวิธีการเขียนระบบหนึ่งไปสู่อีกระบบหนึ่งอย่างมีหลักการ เมื่อปริวรรตแล้วสามารถเขียนกลับเป็นอักษรและอักษรวิธีเดิมได้ (กรรณิการ์ วัฒนเกษม, 2557) การปริวรรตอักษรโบราณเป็นอักษรไทยปัจจุบันประกอบด้วย 2 ขั้นตอนได้แก่ (1) ขั้นตอนการทำ “คำปริวรรต” หมายถึง การถอดรูปอักษรโบราณเป็นอักษรไทยปัจจุบันให้ตรงกับเอกสารต้นฉบับแบบตัวต่อตัว โดยมีหลักเกณฑ์คือคงรูปแบบการประสมอักษรแบบโบราณ หรือ อักษรวิธีการเขียนเหมือนเอกสารต้นฉบับทุกประการ ห้ามเพิ่มเติมพยัญชนะ สระ เครื่องหมายใด ๆ แม้จะพิจารณาแล้วมีความเห็นว่าการใช้พยัญชนะ สระ เครื่องหมาย หรือคำศัพท์เขียนสะกดผิด หรือใช้ผิด ต้องถอดไปตามที่ผิด และหากจะแก้ไขให้ถูกต้องจะต้องทำเป็นเชิงอรรถ หรือคำอธิบายไว้ในที่อื่น หากรูปอักษรลบเลือน แตกหักไม่สมบูรณ์ แต่พอที่จะเดาหรือสันนิษฐานได้จะต้องใส่ตัวอักษรนั้นไว้ในเครื่องหมายวงเล็บ (...) ข้อความใดอ่านไม่ได้ แต่สันนิษฐานได้จากคำที่อยู่ข้าง ๆ มาประกอบ ต้องใส่ข้อความหรือคำนั้นไว้ในวงเล็บเหลี่ยม [...] และตัวอักษรที่อ่านไม่ได้เลยให้ใส่เครื่องหมายจุดไว้ และ (2) ขั้นตอนการทำ “คำอ่าน” ได้แก่ การนำเอาคำปริวรรตไปทำเป็นคำอ่าน โดยคำอ่าน หมายถึง การปรับเปลี่ยนอักษรวิธีการประสมอักษรแบบโบราณให้เป็นแบบอักษรไทยปัจจุบัน หรือสะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน แต่ ชื่อบุคคล ชื่อสถานที่ ชื่อเมือง จะยังคงรูปแบบการเขียนสะกดคำตามรูปแบบเดิมและเขียนเชิงอรรถอธิบายเอาไว้ ทั้งนี้ สำหรับวรรณกรรมไทยที่มีวิธีเขียนใกล้เคียงกับอักษรวิธีอักษรไทยปัจจุบันค่อนข้างมากอาจลดขั้นตอนโดยการตัดขั้นตอนการทำคำปริวรรตออกไป (พิมพ์พรรณ ไพบูล์หวังเจริญ, 2542)

การปริวรรตเอกสารโบราณแบ่งได้เป็น 3 รูปแบบ ได้แก่ (1) การปริวรรตถ่ายถอดรูปอักษรตัวต่อตัวตามต้นฉบับโดยไม่คำนึงถึงเสียงอ่าน วัตถุประสงค์เพื่อรักษารูปแบบการประสมอักษรตามที่ปรากฏในเอกสารโบราณเอาไว้ สำหรับการศึกษาด้านอักษรและภาษาที่ใช้บันทึกในเอกสารโบราณเป็นหลัก มักใช้กับการปริวรรตเอกสารโบราณประเภทจารึก (2) การปริวรรตแบบวรรณกรรมท้องถิ่น หมายถึง การปริวรรตโดยสะกดคำอ่านตามเสียงของภาษาที่บันทึกในเอกสารต้นฉบับ วัตถุประสงค์เพื่อให้อ่านออกเสียงเป็นภาษาถิ่นและการศึกษาทางด้านวรรณกรรมภาษาถิ่น แต่การสะกดคำให้อ่านออกเสียงตรงกับสำเนียงหรือเสียงพูดภาษาถิ่นทำได้ยาก และคนที่ไม่รู้ภาษาถิ่นมักอ่านไม่เข้าใจ และ (3) การปริวรรตแบบวรรณกรรมท้องถิ่นให้แพร่หลายออกไปเป็นการปริวรรตโดยทำเป็นคำอ่านแบบแปล หรือตีความจากภาษาถิ่นโบราณเป็นภาษาไทยปัจจุบัน ไม่คำนึงเสียงในภาษาถิ่นและอักษรวิธีการประสมอักษรตามเอกสารต้นฉบับ ซึ่งเป็นรูปแบบคำอ่านที่ทุกคนทั่วไปอ่านเข้าใจง่าย (ประเสริฐ ฒ นคร, 2541) จะเห็นได้ว่าการปริวรรตเอกสารโบราณในทุกชั้นตอนนั้น จำเป็นต้องอาศัยผู้ที่มีความเชี่ยวชาญรวมถึงมีประสบการณ์ทั้งด้านการอ่านและการแปลอักษรและภาษาโบราณมาเป็นระยะเวลายาวนานในหลายหมวดมาเป็นผู้ดำเนินการ การปริวรรตเอกสารโบราณ โดยมนุษย์จึงเป็นงานที่มีค่าใช้จ่ายสูงทั้งแรงงานและเวลา จึงมีความพยายามในการพัฒนาระบบปริวรรตเอกสารโบราณด้วยเครื่อง โดยใช้เทคนิควิธีการแปลภาษาด้วยเครื่องมาช่วยให้การปริวรรตเอกสารโบราณสามารถทำได้รวดเร็วขึ้น ลดขั้นตอนการทำงานให้น้อยลง รวมถึงอำนวยความสะดวกแก่บุคคลทั่วไปสามารถนำเอาเอกสารโบราณไปศึกษาค้นคว้าทำได้ง่ายและสะดวกยิ่งขึ้น

การแปลภาษาด้วยเครื่อง (Machine Translation: MT) เป็นส่วนหนึ่งของงานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นกระบวนการแปลภาษามนุษย์หรือภาษาธรรมชาติจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่งแบบอัตโนมัติด้วยเครื่องคอมพิวเตอร์ มีประโยชน์ช่วยให้มนุษย์สื่อสารกันได้โดยไม่มีข้อจำกัดเรื่องภาษาที่แตกต่างกัน รวมถึงการลดต้นทุนการแปลภาษาให้ลดลง เทคนิควิธีที่นำมาใช้ในการแปลภาษาด้วยเครื่องโดยทั่วไป แบ่งได้เป็น 3 กลุ่ม ได้แก่ (1) การใช้ความรู้ทางภาษาศาสตร์มาสร้างเป็นกฎในการแปลภาษา (Rule-based Machine Translation: RBMT) คู่กับพจนานุกรมสองภาษา ซึ่งมีข้อจำกัดคือ ภาษาแต่ละภาษามีไวยากรณ์จำนวนมาก จึงยากต่อการเขียนกฎให้ครอบคลุมได้ทั้งหมด และพจนานุกรมมีคำศัพท์ไม่ครอบคลุมคำศัพท์ในประโยคที่ต้องแปล (2) การใช้คลังคำเป็นฐานความรู้ในการแปล (Corpus-based Machine Translation: CMT) เป็นวิธีการแปลภาษาด้วยเครื่องแบบอิงตัวอย่างจากคลังคำที่เทียบบท (Parallel Corpus) ซึ่งเป็นคลังคำคู่ประโยคระหว่างภาษาต้นฉบับและประโยคภาษาเป้าหมาย ร่วมกับการใช้วิธีทางสถิติมาคำนวณหาค่าความเป็นไปได้ของคำแปล หรือการนำมาฝึกให้คอมพิวเตอร์เรียนรู้จากคลังคำเพื่อสร้างตัวแบบการแปลภาษา (Machine Learning Model) ซึ่งเป็นวิธีที่ต้องใช้คลังคำเทียบบทที่มีขนาดใหญ่ และ (3) การแปลแบบผสมวิธีการใช้กฎไวยากรณ์ทางภาษาและคลังคำ (Hybrid Machine Translation) (Kituku, Muchemi, & Nganga, 2016)

เทคนิควิธีการแปลภาษาด้วยเครื่องแต่ละแบบได้ถูกนำไปใช้กับงานด้านการแปลภาษาหลากหลายภาษา รวมถึงงานด้านปริวรรตเอกสารโบราณของไทยด้วย โดยในช่วงระยะเวลากว่า 10 ปี ได้มีความพยายามนำเสนอเทคนิควิธีสำหรับปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรไทยโบราณด้วยเครื่อง อาทิ การแปลงอักษรธรรมอีสานเป็นคำอ่านภาษาไทยถิ่นอีสานและสัทอักษรโดยใช้พจนานุกรมร่วมกับกฎ (Rule) (นงคินุช ไพบุลย์, 2552) การแปลงอักษรธรรมอีสานให้เป็นภาษาไทยโดยใช้ไวยากรณ์ไม่พึ่งบริบท (Context Free Grammar: CFG) และพจนานุกรมอักษรธรรมอีสาน-ไทย (โสภณ พรหมโสดา, 2553) และงานอื่น ๆ ที่เกี่ยวข้องกับการนำเสนอแนวคิด หรือขั้นตอนวิธีสำหรับนำมาใช้ในกระบวนการปริวรรตเอกสารโบราณด้วยเครื่อง อาทิ การวิเคราะห์โครงสร้างคำของภาษาล้านนาโดยใช้เทคนิคการทำเหมืองกราฟสำหรับใช้เป็นตัวแทนของคำภาษาล้านนา (ณัฐธิดา สุรเดช, 2554) การพัฒนาโปรแกรมพจนานุกรมอิเล็กทรอนิกส์ล้านนา-ไทย (พิเชฐ ตันดิพัฒน์ภิญโญ, 2551; นิกร ยาทรม, 2554) และการศึกษาพัฒนาระบบการแบ่งคำอักษรธรรมอีสานเพื่อแก้ปัญหาคำที่ไม่ชัดเจนของคำ และปัญหาคำที่ไม่รู้จักในข้อความอักษรธรรมอีสานที่ไม่มีการเว้นวรรคและความคลุมเครือ (Somsap & Seresangtakul, 2020) โดยพัฒนาตัวแบบการจัดกลุ่มอักษรธรรมอีสาน (Isam Dharma Character Cluster: IDCC) ที่ใช้สถิติเป็นฐานร่วมกับวิธีการรู้จำชื่อเฉพาะ (Named Entity Recognition: NER) เป็นต้น

เทคนิควิธีที่นำมาใช้ในการพัฒนาระบบปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรไทยโบราณด้วยเครื่องเกือบทั้งหมดจะใช้พจนานุกรม ซึ่งมีรายละเอียดเกี่ยวกับคำศัพท์และกฎไวยากรณ์ของภาษามาเป็นฐานความรู้ โดยมีข้อจำกัดคือ อาจเกิดความผิดพลาดจากกรณีมีความกำกวมเกิดขึ้นในขั้นตอนการวิเคราะห์อักษรถิ่นและภาษาที่เขียนในเอกสารโบราณต้นฉบับ (Source Language) และในขั้นตอนของการตั้งเคราะห์ภาษาเป้าหมาย (Target Language) หรือคำอ่านอักษรและภาษาไทยปัจจุบัน เนื่องจากฐานความรู้ที่มีอยู่ไม่พอเพียงสำหรับนำมาใช้แก้ปัญหาคำกำกวมที่เกิดจากคำศัพท์อักษรโบราณมีลักษณะเขียนอย่างหนึ่งอ่านอีกอย่างหนึ่งได้ (ต้องอาศัยบริบทในการอ่านแปลความหมาย) โดยวิธีการหนึ่งที่ช่วยแก้ปัญหาดังกล่าว คือ การแปลภาษาด้วยเครื่องแบบอิงตัวอย่างจากคลังคำเทียบบท หรือคลังคำคู่ขนาน 2 ภาษาของอักษรและภาษาโบราณ (ภาษาดั้งเดิม) และ คำอ่านอักษรและภาษาไทยปัจจุบัน (ภาษาเป้าหมาย) โดยใช้วิธีทางสถิติ หรือการนำมาฝึกให้คอมพิวเตอร์เรียนรู้จากคลังคำเพื่อสร้างตัวแบบการแปลภาษา อย่างไรก็ตาม แม้ว่าวิธีนี้จะได้รับการยอมรับว่าเป็นวิธีที่มีประสิทธิภาพสูง แต่เนื่องจากเป็นวิธีที่ต้องใช้คลังคำเทียบบทที่มีขนาดใหญ่มากจึงเป็นเรื่องยากสำหรับเอกสารโบราณ เนื่องจากในปัจจุบันยังไม่มีคลังคำเทียบบทของอักษรและภาษาไทยโบราณ-คำอ่านอักษรและภาษาไทยปัจจุบันสำหรับนำมาเรียนรู้ด้วยเครื่อง หรือสร้างตัวแบบการแปลภาษาแบบอัตโนมัติ ประกอบกับแบบอักษร (Font) สำหรับพิมพ์ด้วยคอมพิวเตอร์ของอักษรไทยโบราณบางอักษร เช่น อักษรไทยน้อย อักษรฝักขาม หรือ อักษรไทยนิเทศ เป็นอักษรที่ยังไม่มีรหัสยูนิโคด (Unicode) ทำให้ไม่สามารถสร้างคลังคำเทียบบท 2 ภาษาของอักษรไทย

โบราณและอักษรไทยปัจจุบันได้ จากปัญหาดังกล่าว งานวิจัยนี้จึงสนใจที่จะศึกษาและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน เพื่อใช้สำหรับการพัฒนาระบบปริวรรตเอกสาร โบราณด้วยเครื่อง ซึ่งใช้ข้อมูลนำเข้าเป็นคำปริวรรตอักษรไทยปัจจุบันแทนข้อมูลนำเข้าเป็นคำอักษรไทยโบราณ โดยเนื้อหาที่นำมาศึกษาเป็นเนื้อหาด้านตำรายาโบราณอีสานที่บันทึกไว้ในเอกสารโบราณด้วยอักษรไทยน้อย

อักษรไทยน้อยเป็นอักษรที่คนอีสานและคนลาวที่อาศัยอยู่ในบริเวณอาณาจักรล้านช้าง (ปัจจุบันคือบริเวณแถบลุ่มแม่น้ำโขงและสปป.ลาว) ใช้จดบันทึกเรื่องราวทางโลก อาทิ วรรณกรรมและนิทานพื้นบ้าน ตำรายา กฎหมาย พงศาวดาร และเอกสารทางราชการ (หนังสือจุ่ม) มาตั้งแต่ พ.ศ. 1890 โดยจะบันทึกเป็นภาษาไทยถิ่นอีสาน เขียนจารตัวอักษรลงในศิลาจารึก ไบลาน หรือกระดาษข่อย ส่วนมากจะพบเขียนจารลงในไบลาน (ธวัช ปุณ โณทก, 2549) คนไทยถิ่นอีสานใช้อักษรไทยน้อยบันทึกเรื่องราวและหนังสือราชการเรื่อยมาจนกระทั่งเมื่อมีการจัดการเรียนการสอนในระบบโรงเรียนตามพระราชบัญญัติประถมศึกษา พ.ศ. 2465 คนอีสานเริ่มเรียนอักษรไทยปัจจุบันมากขึ้น ทำให้การบันทึกเรื่องราวต่าง ๆ ด้วยอักษรไทยน้อยลดน้อยลงจนกระทั่งเลิกใช้ไปในที่สุด (ธวัช ปุณ โณทก, 2549)

อักษรไทยน้อยประกอบด้วยพยัญชนะเดี่ยวประมาณ 24 - 27 รูป พยัญชนะประสม (พยัญชนะ 2 ตัวเขียนติดกัน) 8 รูป สระ 23 รูป ไม่มีวรรณยุกต์สำหรับกำกับ โทณเสียงสูงต่ำ อักษรวิธี (Orthography) หรือวิธีประสมอักษรให้เป็นคำของอักษรไทยน้อยส่วนใหญ่จะคล้ายกับอักษรไทยคือ วางพยัญชนะต้นและตัวสะกดไว้บนบรรทัด เขียนสระไว้รอบพยัญชนะต้น (หน้า หลัง บน ล่าง) แต่จะมีการใช้รูปอักษรและอักษรวิธีของอักษรธรรมมาปะปนบางส่วน เช่น การใช้รูปพยัญชนะตัวเชิงหรือตัวเฟื้อง “ย” ของอักษรธรรมอีสานเป็นสระเอียสำหรับสร้างคำที่ประสมด้วยสระเอียแบบไม่มีตัวสะกด หรือคำที่มีพยัญชนะควบกล้ำ “ล” จะใช้รูปตัวเชิง ล อักษรธรรมอีสานเขียนไว้ได้ พยัญชนะตัวที่ไปควบกล้ำ หรือใช้รูปพยัญชนะรูปตัวเชิงของอักษรธรรมอีสานบางตัวมาใช้เป็นพยัญชนะตัวสะกด เป็นต้น (ธวัช ปุณ โณทก, 2549; กมล แก้วปลั่ง, 2556)

การปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรไทยน้อยนั้น ผู้ปริวรรตจะต้องมีความรู้ทักษะและความเชี่ยวชาญที่สำคัญ ได้แก่ ขั้นตอนการทำคำปริวรรตจะต้องรู้จักรูปอักษร อักษรวิธีของอักษรไทยน้อยและอักษรธรรมอีสาน ดังนั้นความยากของขั้นตอนนี้ ประการแรกเกิดจากความหลากหลายของรูปอักษรซึ่งเป็นลายมือเขียนที่มีลีลาการเขียนแตกต่างกันในผู้เขียนแต่ละคน บางลายมือนั้นอ่านง่าย แต่บางลายมือนั้นอ่านยาก โดยจะส่งผลต่อการระบุรูปอักษรแต่ละตัวให้ถูกต้องและความยากประการที่สอง เกิดจากการสร้างคำที่ใช้วิธีประสมอักษรด้วยอักษรวิธีพิเศษเฉพาะบุคคล ซึ่งต่างจากวิธีประสมอักษรปกติ เช่น การใช้พยัญชนะตัวสะกดของคำพยางค์หน้าเป็นพยัญชนะต้นของคำพยางค์หลัง (ใช้พยัญชนะร่วมกัน) การสร้างกฎเกณฑ์วิธีประสมอักษรขึ้นมา

ใหม่โดยการเลียนแบบวิธีประสมอักษรในภาษาบาลี หรือ การใช้สัญลักษณ์แทนคำหรือตัวอักษร (สุภณ สมจิตศรีปัญญา, 2525; จันทน์แดง คำลือหาญ, 2535; กิตติวัฒน์ สัตนาโค, 2542; เอมอร เขาวนัสวาน, 2558) เป็นต้น และขั้นตอนการทำคำอ่านจะต้องมีทักษะคือ มีความรู้ภาษาไทยถิ่นอีสาน โบราณ และความรู้เกี่ยวกับเนื้อหาของเรื่องที่อ่าน เนื่องจากคำปวิรรตของอักษรไทยน้อยมีลักษณะ “เขียนอย่างหนึ่งแต่อ่านอีกอย่างหนึ่ง” ซึ่งเกิดจากจำนวนตัวอักษรของอักษรไทยน้อยมีน้อยกว่า อักษรไทยปัจจุบัน และไม่มีวรรณยุกต์ ทำให้คำปวิรรตของอักษรไทยน้อยบางคำมีคำอ่านได้หลาย แบบหลายความหมาย เช่น “โก” (คำปวิรรต) มีคำอ่านได้ 3 คำอ่าน ได้แก่ “โก” “ไกล” “ไกล้” หรือ “ป้า” (คำปวิรรต) มีคำอ่านได้ 3 คำอ่าน ได้แก่ “ป้า” “ป้า” “ปลา” เป็นต้น ซึ่งคำอ่านที่ถูกต้องคือ คำอ่านที่มีความหมายสอดคล้องทั้งบริบทและเนื้อเรื่อง

ตำราชาวโบราณอีสานบันทึกภูมิปัญญาด้านการดูแลสุขภาพและการรักษาโรคด้วย สมุนไพรของคนอีสานเอาไว้เป็นจำนวนมาก ในขณะที่สมุนไพรไทยได้รับการยอมรับว่าเป็นที่สุด แห่งภูมิปัญญาไทย สามารถนำมาศึกษาและพัฒนาต่อยอดเพิ่มมูลค่าและใช้ประโยชน์ในด้านต่าง ๆ อาทิ ใช้รักษาโรคและสร้างเสริมสุขภาพที่ตีร้อมกับการใช้ยาแผนปัจจุบัน การนำมาพัฒนาทำให้เกิด มูลค่าเพิ่มแก่สินค้าและบริการประเภทต่าง ๆ อาทิ นำไปแปรรูปสำหรับใช้ในอุตสาหกรรมผลิตยา จากสมุนไพร อุตสาหกรรมผลิตเครื่องสำอางค์และผลิตภัณฑ์สปา อุตสาหกรรมผลิตสารสกัดจาก สมุนไพร เป็นต้น ประโยชน์ของสมุนไพรไทยจึงไม่เพียงช่วยเสริมสร้างคุณภาพชีวิตที่ดีของ ประชาชน ลดการนำเข้ายาแผนปัจจุบันจากต่างประเทศ และลดค่าใช้จ่ายด้านสุขภาพของประเทศ ในภาพรวมได้เท่านั้น แต่ยังสามารถส่งออกไปยังต่างประเทศสร้างรายได้ให้กับประเทศอย่าง มหาศาล รวมถึงสามารถใช้เป็นเครื่องมือผลักดันเศรษฐกิจของประเทศให้เจริญเติบโตอย่างยั่งยืน (กองยุทธศาสตร์และแผนงาน, 2561; สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ, 2564) ด้วยประโยชน์และคุณค่าของสมุนไพรไทยดังกล่าว การศึกษาค้นคว้าตำราชาวโบราณอีสานจะช่วย ทำให้เกิดองค์ความรู้ใหม่ ๆ สำหรับนำไปพัฒนาเป็นผลิตภัณฑ์จากสมุนไพรไทยให้มีความ หลากหลายและตอบสนองความต้องการของผู้บริโภคยิ่งขึ้น โดยการพัฒนาระบบปวิรรตตำราชาว โบราณอีสานแบบอัตโนมัติเป็นสิ่งหนึ่งที่จะทำให้เกิดการเข้าถึงและนำองค์ความรู้ที่บันทึกไว้ไปใช้ ประโยชน์ทำได้ง่ายขึ้น

การวิจัยนี้เริ่มด้วยการศึกษาปัญหาการปวิรรตตำราชาวโบราณอีสานที่บันทึกในโบราณด้วย อักษรไทยน้อย เพื่อยืนยันปัญหาของการปวิรรตตำราชาวโบราณอีสาน เพื่อนำมาใช้ในการออกแบบ และพัฒนากรอบการทำงานในการแปลงคำปวิรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่น อีสาน ซึ่งประกอบด้วยกรอบการออกแบบและพัฒนาคลังศัพท์ (Lexicon) คำปวิรรต-คำอ่านภาษาไทย ถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน และการออกแบบออนโทโลยีคำศัพท์ตำราชาวโบราณอีสาน โดยคลังศัพท์จะใช้สำหรับจัดเก็บคำปวิรรต คำอ่านภาษาไทยถิ่นอีสาน และคำอ่านภาษาไทย

มาตรฐาน และออนโทโลยีคำศัพท์ตำรายาโบราณอีสานจะใช้สำหรับจัดเก็บคำศัพท์และบริบทของคำศัพท์นั้น แล้วทำการประเมินประสิทธิภาพกรอบการทำงานที่พัฒนาในเชิงคุณภาพ คือ ความเหมาะสมด้านการออกแบบและพัฒนากรอบการทำงานเมื่อนำไปพัฒนาเป็นระบบปริวรรตตำรายาโบราณอีสานด้วยเครื่อง และประสิทธิภาพเชิงปริมาณคือ ระดับความถูกต้องของผลลัพธ์ คำอ่านภาษาไทยถิ่นอีสานทั้งในระดับคำและระดับประโยค

1.2 วัตถุประสงค์การวิจัย

- 1.2.1 เพื่อศึกษาปัญหาในการปริวรรตตำรายาโบราณอีสานที่บันทึกในเอกสารโบราณด้วยอักษรไทยน้อย
- 1.2.2 เพื่อออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน และออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน
- 1.2.3 เพื่อออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน
- 1.2.4 เพื่อประเมินประสิทธิภาพเชิงคุณภาพและเชิงปริมาณของกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

1.3 คำถามการวิจัย

- 1.3.1 กรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานที่มีความหมายถูกต้องตามบริบทของตำรายาโบราณอีสาน ที่ใช้คลังศัพท์และออนโทโลยีเป็นฐานความรู้ นั้น ควรมีองค์ประกอบและกระบวนการอย่างไร
- 1.3.2 คลังศัพท์และออนโทโลยีที่ใช้สำหรับแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานควรมีโครงสร้างและองค์ประกอบอย่างไร

1.4 ขอบเขตการวิจัย

การวิจัยนี้ เป็นวิจัยเพื่อการออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน กำหนดขอบเขตในการวิจัยนี้

- 1.4.1 ขอบเขตด้านเนื้อหา ศึกษาเฉพาะหมวดตำรายาโบราณอีสานที่ปริวรรตมาจากตำรายาโบราณอีสานที่บันทึกในโบราณด้วยอักษรไทยน้อย มีสำเนาภาพดิจิทัลของตำรายาดั้งฉบับ และเป็นตำรายฉบับที่สามารถติดต่อกับผู้ปริวรรตได้เท่านั้น

- 1.4.2 ข้อมูลสำหรับทดลองและทดสอบประสิทธิภาพของกรอบการทำงาน

1.4.2.1 เป็นคำรับยาที่ใช้สมุนไพรในการรักษาเท่านั้น ไม่รวมคำรับยาที่ใช้วิธีรักษาแบบประสมประสานร่วมกับ บทสวด คาถา ทั้งแบบ ภาษาไทยถิ่นอีสาน และภาษาบาลี

1.4.2.2 ข้อมูลสำหรับทดสอบประสิทธิภาพของกรอบการทำงานที่พัฒนาในงานวิจัยนี้ สร้างจากคำอ่านตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ด้วยวิธีแปลงคำอ่านย้อนกลับไปเป็นคำปริวรรต ตามหลักการปริวรรตอักษรไทยน้อยแบบถ่ายถอดตรงตัว โดยละเครื่องหมายที่ใช้สำหรับบอกวิธีประสมอักษรตามเอกสารต้นฉบับ ดังนี้

- 1) ละเครื่องหมาย (T) สำหรับบอกวิธีประสมอักษรด้วยไม้กุ่ม
- 2) ละเครื่องหมาย (-) สำหรับบอกว่ารูปอักษรต้นฉบับเป็นพยัญชนะประสม
- 3) ละเครื่องหมายพินทุ (.) เพื่อให้ระบุว่าอักษรต้นฉบับเป็นพยัญชนะควบกล้ำที่เขียนไว้ได้พยัญชนะต้น

1.4.2.3 ตัวอักษรในเอกสารต้นฉบับที่เขียนผิดแล้วขีดฆ่าเขียนใหม่ ตัวอักษรที่ถูกขีดฆ่า จะไม่ถูกถอดออกมา หรือกรณีเขียนพยัญชนะตัวเดิมหรือคำเดิมซ้ำกันจะถอดออกมาเป็นคำปริวรรตเพียงครั้งเดียว และหากสะกดคำผิดโดยไม่มีกรแก้ไขใด ๆ อักษรนั้นจะยังคงถูกถอดออกมาตามการเขียนในเอกสารต้นฉบับ

1.4.2.4 ข้อมูลตำรายาโบราณอีสานที่นำมาใช้สร้างชุดทดสอบประสิทธิภาพรวบรวมจากตำรายาโบราณอีสานคนละชุดกับที่นำมาสร้างคลังศัพท์และออนโทโลยี

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้กรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

1.5.2 ได้กระบวนการและขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

1.5.3 ได้คลังศัพท์คำปริวรรตของอักษรไทยน้อย-คำอ่านภาษาไทยถิ่นอีสาน-ภาษาไทยมาตรฐาน และออนโทโลยีคำศัพท์ตำรายาโบราณอีสานที่สามารถใช้เป็นฐานความรู้สำหรับการปริวรรตตำรายาโบราณอีสานโดยเฉพาะ

1.5.4 ได้ทราบลักษณะของอักษรไทยน้อยและอักษรวิธีของอักษรไทยน้อยที่ส่งผลกระทบต่อประสิทธิภาพของระบบปริวรรตตำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อย

1.5.5 ได้ระบบต้นแบบสำหรับนำไปใช้ในงานด้านการพัฒนาระบบปริวรรตตำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยด้วยเครื่อง

1.6 คำอธิบายศัพท์

คำอธิบายศัพท์ที่ใช้ในการวิจัยนี้ มีดังนี้

1.6.1 คำปริวรรตของอักษรไทยน้อย หมายถึง คำศัพท์ภาษาไทยถิ่นอีสาน โบราณที่เขียนด้วยอักษรไทยปัจจุบัน แต่ประสมอักษรตรงกับพยัญชนะหรือสระที่บันทึกไว้ในเอกสารโบราณต้นฉบับ

1.6.2 คำอ่านภาษาไทยถิ่นอีสาน หมายถึง คำภาษาไทยถิ่นอีสานที่สะกดคำด้วยอักษรไทยปัจจุบันตามหลักเกณฑ์การสะกดคำของภาษาไทยมาตรฐาน

1.6.3 คำอ่านภาษาไทยมาตรฐาน หมายถึง คำภาษาไทยมาตรฐานซึ่งมีความหมายเทียบเท่าหรือความหมายตรงกันกับคำอ่านภาษาไทยถิ่นอีสาน

1.6.4 คำราชาโบราณอีสาน หมายถึง หนังสือที่บันทึกความรู้เกี่ยวกับวิธีการรักษาหรืออาการเจ็บป่วยของคนอีสานสมัยโบราณ มีหลายโรค หลายอาการ หลายวิธีการรักษา

1.6.5 คำรับยาอีสาน โบราณ หมายถึง วิธีการรักษาโรค หรืออาการเจ็บป่วยของคนอีสานสมัยโบราณ เฉพาะ โรคใดโรคหนึ่ง หรืออาการใดอาการหนึ่ง

1.6.6 การประสมอักษรแบบโบราณ หมายถึง การประสมพยัญชนะและสระอักษรไทยปัจจุบันเป็นคำ ที่ยังคงรูปแบบการใช้พยัญชนะ สระ ตรงตามการประสมอักษรในเอกสารโบราณต้นฉบับ

บทที่ 2

ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนากรอบการทำงานในแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน สำหรับพัฒนาระบบปริวรรตคำรายอีสานโบราณแบบอัตโนมัติ ซึ่งเป็นระบบแปลภาษาด้วยเครื่องประเภทหนึ่ง การปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้องสำหรับการกำหนดกรอบแนวคิดในงานวิจัย มีดังนี้

2.1 อักษรไทยน้อย

2.1.1 ความเป็นมาของอักษรไทยน้อย

2.1.2 อักษรวิธีอักษรไทยน้อย

2.2 การปริวรรตอักษรไทยน้อย

2.2.1 การปริวรรตอักษรโบราณ

2.2.2 หลักเกณฑ์การปริวรรตเอกสารโบราณ

2.2.3 หลักเกณฑ์และขั้นตอนการปริวรรตอักษรไทยน้อย

2.3 ภาษาไทยถิ่นอีสาน

2.3.1 ลักษณะเฉพาะภาษาไทยถิ่นอีสาน

2.3.2 ไวยากรณ์ภาษาไทยถิ่นอีสาน

2.3.3 การประสมคำภาษาไทยถิ่นอีสาน

2.3.4 ประโยคภาษาไทยถิ่นอีสาน

2.4 ออนโทโลยี

2.4.1 ความหมายและความสำคัญของออนโทโลยี

2.4.2 องค์ประกอบของออนโทโลยี

2.4.3 ประเภทของออนโทโลยี

2.4.4 กระบวนการพัฒนาออนโทโลยี

2.4.5 ภาษาที่ใช้ในการพัฒนาออนโทโลยี

2.4.6 โปรแกรมสร้างออนโทโลยี

2.4.7 การประเมินผลออนโทโลยี

2.5 คลังศัพท์

2.5.1 ความหมาย

- 2.5.2 การรวบรวมคำศัพท์เพื่อสร้างคลังศัพท์
- 2.5.3 ขั้นตอนการสร้างคลังศัพท์
- 2.5.4 คลังศัพท์อิเล็กทรอนิกส์
- 2.6 การประมวลผลภาษาธรรมชาติ
 - 2.6.1 ความหมายของการประมวลผลภาษาธรรมชาติ
 - 2.6.2 กระบวนการทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติ
 - 2.6.3 ขั้นตอนการประมวลผลภาษาธรรมชาติ
- 2.7 การแปลภาษาด้วยเครื่อง
 - 2.7.1 ความหมาย
 - 2.7.2 ประเภทของระบบแปลภาษาด้วยเครื่อง
- 2.8 งานวิจัยที่เกี่ยวข้อง
 - 2.8.1 การแปลภาษาไทยโบราณด้วยเครื่อง
 - 2.8.2 การพัฒนาคลังศัพท์อิเล็กทรอนิกส์สำหรับงานประมวลผลภาษาธรรมชาติ
 - 2.8.3 การใช้ออนโทโลยีในการประมวลผลภาษาธรรมชาติ
- 2.9 กรอบแนวคิดในการวิจัย

2.1 อักษรไทยน้อย

2.1.1 ความเป็นมาของอักษรไทยน้อย

อักษรไทยน้อยเป็นอักษรไทยโบราณอักษรหนึ่งที่คนไทยถิ่นอีสานและคนลาวที่อาศัยในบริเวณแถบลุ่มแม่น้ำโขงใช้จดบันทึกเรื่องราวทางโลก อาทิ วรรณกรรมและนิทานพื้นบ้าน ตำรายา กฎหมาย พงศาวดาร และเอกสารทางราชการ (หนังสือพิมพ์) โดยบันทึกเป็นภาษาไทยถิ่นอีสานหรือภาษาลาว วิธีบันทึกใช้วิธีเขียนจารตัวอักษรลงในวัสดุประเภทต่าง ๆ เช่น ศิลาจารึก ใบลาน หรือ กระดาษ พบมากในการเขียนจารลงในใบลานและสมุดข่อย รูปอักษรไทยน้อยมีพัฒนาการแบ่งได้เป็น 2 ระยะ ได้แก่ ระยะที่ 1 (พ.ศ. 1890 - 1911) พัฒนามาจากรูปอักษรไทยสมัยพระเจ้าลิไท แห่งอาณาจักรสุโขทัย และระยะที่สอง (พ.ศ. 2063 - 2093) พัฒนาจากรูปอักษรฝักขาม ซึ่งแพร่มาจากอาณาจักรล้านนาและเข้าสู่อาณาจักรล้านช้างในสมัยพระเจ้าโพธิ์สาลราช อักษรไทยน้อยถูกใช้บันทึกเรื่องราวและเอกสารทางราชการเรื่อยมาจนกระทั่งมีพระราชบัญญัติประถมศึกษา พ.ศ. 2464 ซึ่งทำให้เกิดการเปลี่ยนแปลงการจัดการเรียนการสอนจากวัดเป็นระบบโรงเรียน คนไทยถิ่นอีสานหันมาเรียนอักษรไทยปัจจุบันมากขึ้น ความสนใจในการเรียนและใช้อักษรไทยน้อยเพื่อจดบันทึกเรื่องราวต่าง ๆ จึงลดน้อยลงจนกระทั่งเลิกใช้ไปในที่สุด (ชวช ปุณโณทก, 2549)

อักษรไทยน้อยที่ใช้เขียนในเอกสารโบราณแต่ละชนิดพบว่าจะมีจำนวนแตกต่างกัน เช่น ในใบจุ่มซึ่งเป็นเอกสารทางราชการจะพบการใช้พยัญชนะเดี่ยว 33 รูป ในขณะที่ในเอกสารโบราณจะพบ 24 - 29 รูป พยัญชนะประสม หรือพยัญชนะที่มีรูปพยัญชนะ 2 ตัวเขียนติดกัน ในใบจุ่มพบ 12 รูป และในโบราณ พบ 8 รูป โดยเอกสารโบราณที่จารด้วยอักษรไทยน้อยในสมัยรัตนโกสินทร์ที่บันทึกเนื้อหาประเภทวรรณกรรมพื้นบ้าน หรือเอกสารทั่วไปไม่นิยมใช้พยัญชนะ ฃ ฅ ฌ ฎ ฏ ฑ ฒ ณ เนื่องจากการประสมอักษรเป็นคำ มุ่งเน้นให้อ่านออกเสียงตรงตามเสียงภาษาพูด จึงไม่เคร่งครัดเรื่องการประสมอักษรเหมือนกับการบันทึกคำภาษาบาลีที่มีเกณฑ์การสะกดคำที่เคร่งครัดมากกว่า (ธวัช ปุณโณทก, 2549) ซึ่งในงานวิจัยนี้เป็นการศึกษาเฉพาะอักษรไทยน้อยที่ใช้บันทึกตำรายาโบราณอีสานในโบราณ ดังนั้น การปริทัศน์วรรณกรรมที่เกี่ยวข้องที่จะนำเสนอในลำดับถัดไป จึงนำเสนอเฉพาะอักษรไทยน้อยที่ใช้บันทึกในโบราณเท่านั้น

2.1.2 อักษรวิธีอักษรไทยน้อย

อักษรวิธี (Orthography) หมายถึง แบบแผนว่าด้วยตัวหนังสือ วิธีเขียน อ่าน และใช้ตัวหนังสือให้ถูกต้องตามความนิยม ประกอบด้วย ลักษณะอักษร วิธีประสมอักษรเป็นคำ และวิธีใช้อักษร (พระยาอุปกิตศิลปสาร (นิ่ม กาญจนาชีวะ), 2544) หรือ ความหมายตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง วิธีเขียนและอ่านหนังสือให้ถูกต้อง กล่าวโดยสรุป อักษรวิธีอักษรไทยน้อย หมายถึง รูปลักษณะอักษรไทยน้อย ระเบียบแบบแผน วิธีการ หรือหลักเกณฑ์ที่ใช้สำหรับการประสมอักษรไทยน้อยเป็นคำ รายละเอียดมีดังนี้

2.1.2.1 รูปอักษรไทยน้อย แบ่งเป็น 4 ประเภท (ธวัช ปุณโณทก, 2549) ดังนี้

1) พยัญชนะ พยัญชนะอักษรไทยน้อยแบ่งเป็น 2 ประเภท คือ รูปพยัญชนะเดี่ยว และรูปพยัญชนะประสมและตัวควบกล้ำ โดยรูปพยัญชนะเดี่ยวใช้เป็นพยัญชนะต้นและตัวสะกด มี 24 รูป โดยพยัญชนะอักษรไทยน้อยที่ไม่นิยมใช้เมื่อเทียบกับอักษรไทยปัจจุบัน ได้แก่ พยัญชนะ ฃ ฅ ฌ ฎ ฏ ฑ ฒ ณ ด ต ถ ท น บ ป ผ ฝ พ ฟ ภ ม ย ร ฤ ล ฦ ว ศ ษ ส ห ฬ อ ฮ ฯ ะ ั และ า และรูปพยัญชนะประสม ซึ่งเป็นรูปอักษรที่มีลักษณะเป็นตัวอักษร 2 ตัว เขียนเชื่อมติดกัน โดยพยัญชนะตัวแรกเป็นอักษรรูปตัวเต็ม และพยัญชนะตัวที่ 2 หรือตัวตาม เป็นรูปอักษรครึ่งตัว มี 8 รูป ได้แก่ อย หน หม หล คน พย และ ขน ยกเว้นรูปพยัญชนะ อย จะเป็นรูป ย หางยาว ส่วนพยัญชนะประสมหรือพยัญชนะต้นที่มีตัวตามหรือตัวควบกล้ำเป็นพยัญชนะ ล จะเขียนพยัญชนะ ล รูปตัวเชิงอักษรธรรมอีสานไว้ได้ พยัญชนะต้น ทั้งนี้ในการปริวรรตตำรายาโบราณอีสานที่บันทึกในโบราณด้วยอักษรไทยน้อยจะพบรูปพยัญชนะอักษรไทยน้อยที่ใช้เขียนเพิ่มเติม คือ รูปพยัญชนะเดี่ยว ฮ และพยัญชนะควบกล้ำ กล ตัวอย่างรูปพยัญชนะทั้ง 2 ประเภท ดังรูปที่ 2.1 และ 2.2

ก ฅ ด ฎ จ ฌ ๑ ๓ ๓ ๓ ๓
 ก ข ก ง จ ช ด ต ถ ท น
 บ ป ฝ ฝ พ ฟ ม ย ล ๑ ๑
 บ ป ฝ ฝ พ ฟ ม ย ล ว ส
 ท อ ๑
 ห อ ๑

รูปที่ 2.1 พยัญชนะเดี่ยวอักษรไทยน้อย

ฅ ๓ ๓ ๓ ๓ ๑ ๓ ๓ ๓
 อย หน หม หล คน พย ขน หย กล

รูปที่ 2.2 พยัญชนะประสมอักษรไทยน้อย

2) สระ สระอักษรไทยน้อยมี 23 รูป รูปสระส่วนใหญ่คล้ายคลึงกับรูปสระในอักษรไทยปัจจุบัน โดยส่วนที่แตกต่างคือ สระอักษรไทยน้อยบางตัวจะมีรูปสระมากกว่า 1 รูป ดังนี้ “สระเอีย” มี 2 รูป ได้แก่ รูปสระเอียสำหรับการประสมคำแบบมีตัวสะกด และรูปสระเอียสำหรับการประสมคำแบบไม่มีตัวสะกด “สระออ” มี 3 รูป ได้แก่ รูปสระออสำหรับการประสมคำแบบมีตัวสะกด รูปเครื่องหมายนิคหิต (◌ิ) สำหรับการประสมคำแบบไม่มีตัวสะกด และรูปพยัญชนะรูปตัวเชิงหรือตัวเพ็อง ข อักษรธรรมอีสาน สำหรับการประสมคำที่มีตัวสะกดเป็นแม่เกย (ย) “สระไอ” มี 1 รูป ใช้เป็นทั้งสระไอไม้มีลายและสระไอไม้มีวน ตัวอย่างรูปสระอักษรไทยน้อยแสดงดังรูปที่ 2.3

อะ	อา	ไม้หันอากาศ	อิ	อี	อี้	อุ	อู
เอ	แเอ	โอ	เอาะ	ออ	อัว	เอีย	
เออ	เออ	อำ	ไอ	เอา	ไม้มก/กุ่ม	นิคหิต	หอยหขาดน้ำ

รูปที่ 2.3 รูปสระอักษรไทยน้อย

โทนเสียง

- 3) วรรณยุกต์ อักษรไทยน้อยไม่ใช้วรรณยุกต์เป็นเครื่องหมายบอกระดับ
- 4) ตัวเลข รูปตัวเลขอักษรไทยน้อยมี 10 รูป ประกอบด้วย ตัวเลข 0 - 9

ตัวอย่างดังรูปที่ 2.4

๑	๒	๓	๔	๕	๖	๗	๘	๙	๐
๑	๒	๓	๔	๕	๖	๗	๘	๙	๐

รูปที่ 2.4 ตัวเลขอักษรไทยน้อย

2.1.2.2 วิธีการประสมอักษรไทยน้อยเป็นคำ

หลักเกณฑ์สำหรับประสมอักษรไทยน้อยเป็นคำส่วนใหญ่เป็นแบบเดียวกับอักษรไทยปัจจุบัน คือ วางพยัญชนะต้น ตัวตาม และตัวสะกด ไว้บนบรรทัดระดับเดียวกัน และวางสระไว้รอบพยัญชนะต้น (หน้า หลัง บน ล่าง) โดยมีส่วนที่แตกต่างจากวิธีประสมอักษรไทยปัจจุบัน สรุปได้ดังนี้ (กมล แก้วปลั่ง, 2556)

- 1) พยัญชนะตัวสะกดของอักษรไทยน้อยมี 8 รูป แบ่งตามมาตรา ตัวสะกด 8 มาตรา มาตราละ 1 รูป 1 เสียง ได้แก่ แม่กง ใช้รูปพยัญชนะ ' (ง) แม่กน ใช้รูปพยัญชนะ o (น) แม่กม ใช้รูปพยัญชนะ , (ม) แม่เกย ใช้รูปพยัญชนะรูป p (ย) แม่เกว ใช้รูปพยัญชนะ ; (ว) แม่กก ใช้รูปพยัญชนะ d (ก) แม่กค ใช้รูปพยัญชนะ f (ค) และ แม่กบ ใช้รูปพยัญชนะ [(บ)

- 2) คำที่ประสมด้วยสระโอะลดรูป สระอัว หรือ สระเอา ใช้ไม้กงหรือไม้ก้ม เขียนไว้บนพยัญชนะต้น เช่น 8To (คน) oTd (นก) g-Tk (เขา) เป็นต้น
- 3) คำที่ประสมด้วยสระออ แบ่งเป็นคำที่ประสมกับสระออแบบไม่มีตัวสะกด ใช้เครื่องหมายนิคหิตเขียนไว้บนพยัญชนะต้น เช่น rY (พ่อ) [Y (บ) เป็นต้น คำที่ประสมด้วยสระออที่มีตัวสะกดแม่เกย หรือ ย ใช้รูปหอยหยาดน้ำเขียนต่อท้ายพยัญชนะต้นเพื่อใช้เป็นสระออและตัวสะกด “ย” เช่น oZ (น้อย) 8Z (คอย) และคำที่ประสมด้วยสระออที่มีตัวสะกดมาตราอื่น ใช้รูปสระออไทยน้อย เช่น [vd (บอก) <v[(หมอบ) เป็นต้น
- 4) คำที่มีพยัญชนะต้นเป็นพยัญชนะประสม ใช้รูปพยัญชนะประสมเป็นพยัญชนะต้น เช่น <kd (หมา) OZ (หน้อย) สำหรับคำที่ประสมด้วยพยัญชนะควบกล้ำ “ล” จะเขียนพยัญชนะควบกล้ำไว้ได้พยัญชนะต้น เช่น d;k; (กล่าว) เป็นต้น
- 5) คำที่ประสมด้วยสระเอียไม่มีตัวสะกด ใช้รูปสระเอียแบบเต็มรูป เช่น gly\ (เสียด) g-y\ (เจียด) เป็นต้น หากเป็นคำที่ประสมด้วยสระเอียมีตัวสะกดใช้พยัญชนะ ย อักษรไทยน้อยแทนสระเอีย เช่น ;Pd (เวียด) OP' (เจียง) เป็นต้น
- 6) คำที่ประสมด้วยสระเอะ หรือสระเอะ มีตัวสะกด ใช้ไม้อันหรือไม้ขีด เช่น gxyo (เป็น) c-y' (แข็ง) เป็นต้น
- 7) คำที่ออกเสียง อย นำ ใช้รูปหอยหยาดน้ำ หรือ ย หางยาว เช่น ?kd (อยาก) ?6 (อยู่) เป็นต้น

ตัวอย่างการประสมอักษรไทยน้อยเป็นคำ ดังแสดงในรูปที่ 2.5

พยัญชนะต้น ตัวสะกด วางไว้บนบรรทัดเดียวกัน เขียนสระไว้รอบพยัญชนะ

	(เหลือ)		(ห่อ)	

สระออมี 3 รูป

รูปสระเอียมี 2 รูป สำหรับประสมคำที่มีตัวสะกดและไม่มีตัวสะกด

รูปที่ 2.5 ตัวอย่างวิธีการประสมอักษรไทยน้อยเป็นคำ

2.2 การปริวรรตอักษรไทยน้อย

2.2.1 การปริวรรตอักษรโบราณ

การปริวรรตอักษรโบราณ (Transliterations) หมายถึง การถ่ายถอดหรือเปลี่ยนรูปอักษรและอักษรวิธีแบบหนึ่งมาเป็นตัวอักษรและอักษรวิธีหนึ่งอย่างมีหลักเกณฑ์ (กรรณิการ์ วิมลเกษม, 2548) จุดมุ่งหมายของการปริวรรตคือ เพื่อให้ได้ความหมายถูกต้องตรงกับความจริงที่ปรากฏในเอกสารโบราณให้มากที่สุด รวมทั้งทำให้ผู้อื่นสามารถอ่านเนื้อหาที่บันทึกในเอกสารโบราณนั้นได้ (พิมพ์พรรณ ไพบูลหวังเจริญ, 2542) โดยในการปริวรรตอักษรไทยโบราณเป็นอักษรไทยปัจจุบันแบ่งเป็น 3 รูปแบบ (ประเสริฐ ฤ นคร, 2541) ดังนี้

2.2.1.1 ปริวรรตตรงตามรูปอักษรโบราณแบบตัวต่อตัวโดยไม่คำนึงถึงเสียงอ่าน เป็นรูปแบบที่รักษารูปแบบการประสมอักษรแบบเดิมเอาไว้ทุกประการ สามารถเขียนกลับเป็นอักษรและอักษรวิธีเดิมได้ นิยมใช้สำหรับการปริวรรตเอกสารโบราณประเภทจารึก

2.2.1.2 การปริวรรตอักษรตามเสียงในภาษาถิ่น หมายถึง การถ่ายถอดจากอักษรโบราณมาเป็นอักษรไทยปัจจุบันตามเสียงในภาษาถิ่น ซึ่งอาจมีการเติมวรรณยุกต์เพื่อให้อ่านออกเสียงได้ความหมายตรงกับเสียงภาษาถิ่นนั้น ๆ ใช้สำหรับการปริวรรตวรรณกรรมท้องถิ่น สำหรับให้คนในท้องถิ่นนำไปศึกษา แต่คนท้องถิ่นอื่นที่ไม่มีความรู้ในภาษาถิ่นนั้นมักอ่านไม่ค่อยเข้าใจ

2.2.1.3 การปริวรรตเพื่อการเผยแพร่ หมายถึง การปริวรรตหรือถ่ายถอดตัวอักษรโบราณไปเป็นอักษรและภาษาไทยมาตรฐาน โดยไม่คำนึงเสียงในภาษาถิ่นและอักษรวิธีการเขียนในเอกสารโบราณต้นฉบับ หลักในการปริวรรตคือ ถ้ายกคำในภาษาถิ่นใดตรงกับภาษาไทยมาตรฐานให้ใช้คำภาษาไทยมาตรฐาน ยกเว้นคำภาษาถิ่นที่ไม่มีใช้ในภาษาไทยมาตรฐานจะเขียนตามเสียงภาษาถิ่นหรือตามการเขียนในเอกสารโบราณต้นฉบับ ตัวอย่างการปริวรรตถ่ายถอดจากภาษาถิ่นเป็นภาษาไทยมาตรฐาน เช่น “จ้าง” ในภาษาล้านนา หรือ “ช่าง” ในภาษาไทยถิ่นอีสาน จะถูกถ่ายถอดเป็นคำภาษาไทยภาคกลางคือ “ช่าง” เป็นต้น

2.2.2 หลักเกณฑ์การปริวรรตเอกสารโบราณ

หลักเกณฑ์ที่ใช้ในการปริวรรตหรือการถ่ายถอดอักษรโบราณเป็นอักษรไทยปัจจุบันโดยทั่วไป (พิมพ์พรรณ ไพบูลหวังเจริญ, 2542) มีดังนี้

2.2.2.1 การปริวรรตแบบรักษารูปแบบการประสมอักษรตามแบบโบราณ จะต้องคำนึงถึงความถูกต้องของรูปอักษรและภาษาที่เขียนในเอกสารโบราณต้นฉบับ ดังนี้

1) รักษาการใช้พยัญชนะ สระ และเครื่องหมายต่าง ๆ รวมถึงตำแหน่งการวางตัวอักษรให้ตรงตามเอกสารต้นฉบับทุกประการ แม้ว่าจะเป็นการสะกดผิด หรือใช้คำผิด หากต้องการแก้ไขจะเขียนไว้ในเชิงอรรถ หรืออธิบายไว้ที่ตำแหน่งอื่น

2) หากเอกสารชำรุด ข้อความหลงเหลือ ไม่ชัดเจน ฉีกขาด หรือชำรุด บางส่วนแต่พอจะคาดเดาได้ว่าเป็นอักษรอะไรจะใช้เครื่องหมายวงเล็บ () กำกับตัวอักษรนั้นไว้ หรือหากข้อความขาดหายไป แต่พอจะคาดเดาได้จากบริบทได้จะใส่ข้อความนั้นไว้ในวงเล็บ [] หรือหากอ่านไม่ได้เลยจะใส่เครื่องหมายจุด (.) เท่ากับจำนวนตัวอักษรที่อ่านไม่ได้

3) เอกสารโบราณที่บันทึกด้วยอักษรไทยเป็นภาษาไทยที่รูปอักษร ส่วนใหญ่ใกล้เคียงกับอักษรไทยปัจจุบัน แต่มีอักษรวิธีประสมอักษรแตกต่างกัน การถ่ายถอดเป็น อักษรปัจจุบันยังคงใช้หลักการแบบเดียวกันกับข้อ 1 – 2

2.2.2.2 การเขียนคำอ่าน หลังจากปริวรรตอักษรโบราณเป็นอักษรไทยปัจจุบัน หรือทำเป็นคำปริวรรตแล้ว คำปริวรรตยังคงยากต่อการอ่าน เนื่องจากการสะกดคำยังคงเป็นแบบ โบราณ (ประสมอักษรเป็นคำด้วยพยัญชนะ สระ ต่างจากปัจจุบัน) ดังนั้นจึงต้องทำ “คำอ่าน” เพื่อให้ผู้อ่านที่ไม่คุ้นเคยกับการอ่านเอกสารโบราณสามารถอ่านเข้าใจได้ง่ายขึ้น โดยมีเกณฑ์ดังนี้

1) การสะกดคำอ่าน ใช้เกณฑ์การสะกดคำตามที่ใช้ในพจนานุกรมฉบับราชบัณฑิตยสถาน

2) ชื่อบุคคล สถานที่ หรือ ชื่อเมือง ให้เขียนคำอ่านโดยคงรูปเดิมไว้เพื่อใช้ศึกษาด้านประวัติความเป็นมาของคำ และเขียนเชิงอรรถอธิบายชื่อปัจจุบันที่ถูกต้องกำกับเอาไว้

3) คำที่อ่านออกเสียงเหมือนปัจจุบัน ให้ปรับวิธีเขียนคำตามอักษรวิธีปัจจุบัน เช่น ตาลาชา – ตารายา ลิตทิกาน - สิทธิการ สีน - สีล เป็นต้น

2.2.3.3 การชำระวรรณกรรมอักษรไทยโบราณที่บันทึกเป็นภาษาไทย ได้แก่ วรรณกรรมที่มีวิธีเขียนใกล้เคียงกับอักษรวิธีอักษรไทยปัจจุบันค่อนข้างมาก สามารถตัดขึ้นตอนในการทำคำปริวรรตออกไปได้ ดังนี้

1) ปรับอักษรวิธีการประสมอักษรจากคำที่แตกต่างกันมาเป็นคำไทยปัจจุบันตามพจนานุกรม เช่น กาน เขียนเป็น กาล การ (ขึ้นอยู่กับบริบท) หรือ ปะมาด เขียนเป็น ประมาท เป็นต้น

2) ถ้อยคำหรือเนื้อความใด ๆ ที่แตกต่างกันให้ใช้หลักภาษาไทยมาประกอบการพิจารณา พร้อมกับค้นหาหลักฐานอ้างอิงอื่นทางประวัติศาสตร์และโบราณคดีมาประกอบก่อน แล้วจึงเลือกใช้คำที่น่าเชื่อถือมากที่สุดและอธิบายไว้ที่เชิงอรรถ

3) ถ้อยคำหรือเนื้อความที่แตกต่างกันในต้นฉบับหลายฉบับ ต้องมีเชิงอรรถอธิบายความแตกต่างของแต่ละฉบับไว้ทั้งหมด และพยายามรักษาอักษรวิธีเดิมในต้นฉบับเอาไว้ให้มากที่สุด

4) คำใดที่ลึกลับที่เกิดจากการคัดลอก ทั้งกรณีแนใจ และไม่แน่ใจว่าเป็นความผิดพลาดจากการคัดลอก สามารถแก้ไขใหม่ให้ถูกต้องได้เลย

5) คำที่ใช้อักษรเก่า ปัจจุบันไม่ใช่แล้ว เช่น ข ค ให้ใช้ตามภาษาไทยปัจจุบัน

6) จัดเรียงลำดับวรรคให้ถูกต้องตามฉันทลักษณ์ของคำประพันธ์

7) ตัดเครื่องหมายที่ไม่มีมีความสำคัญต่อเนื้อความออกไป

8) จบบทใช้เครื่องหมาย ฯ

9) ทำคำอธิบายคำศัพท์ยากเพื่อให้ความรู้แก่ผู้สนใจ

2.2.3 หลักเกณฑ์และขั้นตอนการปริวรรตอักษรไทยน้อย

การปริวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบัน แบ่งเป็น 2 ขั้นตอน ได้แก่ การทำคำปริวรรต และการทำคำอ่าน แต่ละขั้นตอนมีหลักเกณฑ์ในการดำเนินการ (พิมพ์พรรณ ไพบูลย์หวังเจริญ, 2543) ดังนี้

2.2.3.1 ขั้นตอนทำคำปริวรรต หมายถึง ขั้นตอนการถอดอักษรไทยน้อยเป็นอักษรไทยปัจจุบันแบบตรงตัว ตัวต่อตัว โดยรักษารูปแบบการประสมอักษรตามเอกสารต้นฉบับผลลัพธ์ที่ได้เรียกว่า “คำปริวรรต” ซึ่งหลักเกณฑ์การทำคำปริวรรตมีดังนี้

1) พยัญชนะประสมอักษรไทยน้อยเมื่อถอดเป็นอักษรไทยปัจจุบัน จะใช้เครื่องหมายขีดขวาง (-) เขียนไว้บนพยัญชนะทั้งสองตัวนั้น เช่น กน หน หิม อย เป็นต้น

2) พยัญชนะควบกล้ำ และพยัญชนะตัวตาม หรือตัวสะกดอื่น ที่เขียนไว้ได้พยัญชนะต้นตามแบบอักษรวิธีการประสมอักษรเป็นคำของอักษรธรรมอีสาน หรือคำสองพยางค์ที่พยัญชนะต้นของพยางค์ที่สองเขียนได้พยัญชนะตัวสะกดของพยางค์แรกตามวิธีประสมอักษรภาษาบาลี จะใช้เครื่องหมายจุด (.) ไว้ใต้พยัญชนะต้น หรือพยัญชนะตัวสะกดของพยางค์แรก เพื่อแสดงให้เห็นว่าพยัญชนะตัวถัดมาตำแหน่งเดิมถูกเขียนไว้ใต้พยัญชนะก่อนหน้า เช่น กุลาว หลวง ไหล เป็นต้น

3) คำที่ประสมด้วยสระเอียมมีตัวสะกด ที่ใช้รูปตัวเชิง ย อักษรธรรมอีสาน มาเขียนเป็นสระเอีย ให้ใส่จุดไว้ที่พยัญชนะต้น เช่น จยง (เจียง) เป็นต้น

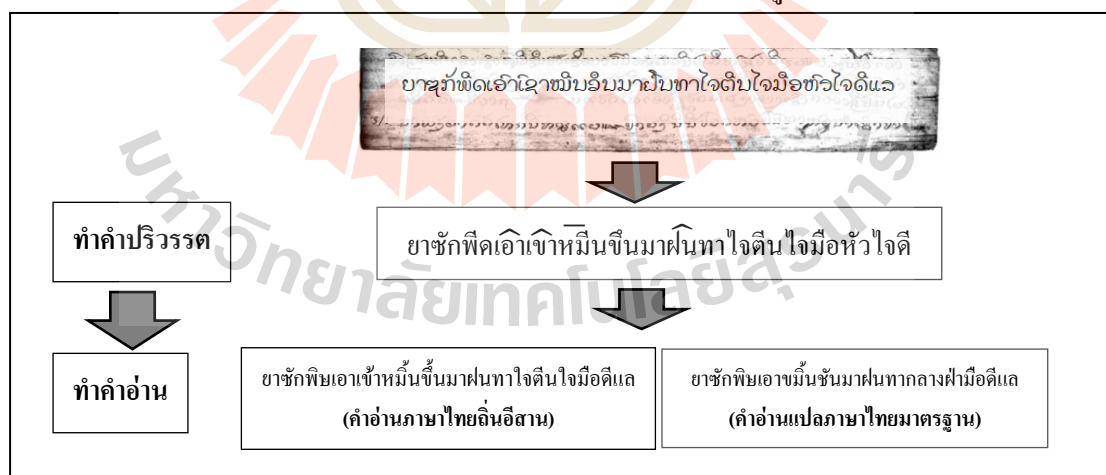
4) คำที่มีไม้ก่ง หรือ ไม้ก๋ม ประสมในคำ เมื่อถอดเป็นอักษรไทยปัจจุบัน จะต้องเขียนไม้ก่ง/ไม้ก๋มไว้ในตำแหน่งเดียวกันกับต้นฉบับ เช่น เรา คน เป็นต้น

5) คำที่เขียนด้วยอักษรวิธีพิเศษ ให้ถ่ายถอดตรงตามที่ปรากฏในต้นฉบับ

2.2.3.1 ขั้นตอนทำคำอ่าน หมายถึง ขั้นตอนการทำคำอ่านของคำปริวรรตแต่ละคำ เพื่อให้อ่านเข้าใจง่ายขึ้น มีหลักเกณฑ์ ดังนี้

- 1) ปรับวิธีการประสมอักษรเป็นคำในแบบโบราณให้เป็นแบบอักษรไทยปัจจุบัน ตามพจนานุกรมฉบับราชบัณฑิตยสถาน ฉบับล่าสุด เช่น “เปือก” เขียนเป็น “เปลือก” “พิด” เขียนเป็น “พिय” เป็นต้น
- 2) ชื่อบุคคล สถานที่ ชื่อเฉพาะ เขียนสะกดคำตามเสียงเดิม เช่น “เมืองชะทอน” “บ้านลากางเมือง” เป็นต้น
- 3) คำภาษาบาลี เขียนตามอักษรวิธีบาลี เช่น “ไชยน์ตะพะวง” เขียนเป็น “ชยตุ ภาว” หรือ “ไชยะมังคะลง” เขียนเป็น “ชยมงคฺล” เป็นต้น
- 4) วัน เดือน ปี ฤกษ์ จะเขียนตามที่ปรากฏ เช่น วัน ๔ เดือนยี่ ปีระวาย ฤกษ์ ๓ ลูก เป็นต้น
- 5) ตัว ฤ ฎ ฏ ในคำปริวรรตเมื่อเขียนเป็นคำอ่านจะเขียนว่า ฤกษ์ ทั้งหมด
- 6) คำที่ออกเสียงเหมือนภาษาไทยมาตรฐาน และเป็นคำที่มีความหมายเดียวกัน แต่เขียนด้วยพยัญชนะต้นแตกต่างกัน จะเขียนตามอักษรวิธีภาษาไทยมาตรฐาน เช่น “ชะบับ/ชาบับ” เขียนเป็น “จบบับ” “ปะหมาน” เขียนเป็น “ประมาณ” “ลินชะเล” เขียนเป็น “ลั่นทะเล” เป็นต้น

ตัวอย่างการทำคำปริวรรตและคำอ่านของอักษรไทยน้อย ดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างขั้นตอนการปริวรรตอักษรไทยน้อย

การศึกษารูปอักษรไทยน้อยและอักษรวิธีการประสมอักษรไทยน้อยเป็นคำ รวมถึงหลักเกณฑ์การปริวรรตอักษรไทยน้อยนั้น มีความสำคัญอย่างยิ่งต่อขั้นตอนการจัดทำคำปริวรรต และการตรวจทานความถูกต้องของคำปริวรรตกับตำรายาโบราณอีสานต้นฉบับที่นำมาศึกษาในงานวิจัยนี้ อย่างไรก็ตาม การจัดเก็บข้อมูลคำปริวรรตในคอมพิวเตอร์จำเป็นต้องมีการปรับเกณฑ์การปริวรรตบางประการเพื่อความสะดวก เช่น การละเครื่องหมายขีดบนที่ใช้สำหรับกำกับพยัญชนะอักษรไทยปัจจุบันที่ถ่ายทอดมาจากพยัญชนะประสมอักษรไทยน้อย และการละไม้ก่งหรือไม้ก้มที่ใช้กำกับการถ่ายทอดคำที่ประสมด้วยสระโอะลครูป เป็นต้น เนื่องจากเครื่องหมายเหล่านี้เป็นเครื่องหมายพิเศษที่ไม่มีให้ใช้บนแป้นพิมพ์

2.3 ภาษาไทยถิ่นอีสาน

การประมวลผลภาษาธรรมชาติเป็นงานที่ต้องอาศัยความรู้ด้านภาษาศาสตร์ และการศึกษาหลักไวยากรณ์ภาษาไทยถิ่นอีสาน เพื่อให้ทราบลักษณะเฉพาะของภาษาไทยถิ่นอีสาน เช่น คำศัพท์และโครงสร้างในการประสมคำ ชนิดของคำ ซึ่งเป็นความรู้สำคัญสำหรับการตัดคำ และการพัฒนาลังศัพท์ รายละเอียดมีดังนี้

2.3.1 ลักษณะเฉพาะภาษาไทยถิ่นอีสาน

ประเทศไทยมีภาษาถิ่นที่ใช้พูดในภูมิภาคต่าง ๆ แบ่งกลุ่มตามภูมิศาสตร์และจำนวนผู้พูดภาษานั้นอาศัยอยู่เป็นส่วนใหญ่ แบ่งได้เป็น 4 กลุ่ม คือ (1) ภาษาไทยถิ่นกลาง คือ ภาษาพูดที่ใช้ในจังหวัดกรุงเทพมหานคร และจังหวัดรอบ ๆ บริเวณพื้นที่ภาคกลางของประเทศ เช่น นครปฐม ปทุมธานี สมุทรปราการ (2) ภาษาไทยถิ่นใต้ ได้แก่ ภาษาที่ใช้ในจังหวัดทางภาคใต้ของประเทศ เช่น สงขลา ภูเก็ต พังงา สุราษฎร์ธานี นครศรีธรรมราช เป็นต้น (3) ภาษาถิ่นอีสาน ได้แก่ ภาษาพูดที่ใช้ในจังหวัดทางภาคตะวันออกเฉียงเหนือของประเทศ เช่น มหาสารคาม ร้อยเอ็ด กาฬสินธุ์ ชัยภูมิ อุดรธานี และ (4) ภาษาไทยถิ่นเหนือ ได้แก่ ภาษาพูดที่ใช้ในจังหวัดทางภาคเหนือของประเทศ เช่น เชียงใหม่ เชียงราย ลำพูน ลำปาง เป็นต้น โดยภาษาถิ่นที่ถูกกำหนดให้เป็นภาษาเพื่อการติดต่อสื่อสารอย่างเป็นทางการคือ ภาษาไทยถิ่นกลางสำเนียงกรุงเทพฯ เรียกว่า ภาษาไทยมาตรฐาน นอกจากนี้ ในแต่ละท้องถิ่นยังมีภาษาย่อย ๆ ที่ใช้เฉพาะกลุ่มอีกเป็นจำนวนมาก แต่จำนวนผู้ใช้ มีไม่มากเมื่อเทียบกับภาษาถิ่นทั้ง 4 ภาษา (วิไลศักดิ์ กิ่งคำ, 2556)

ภาษาไทยถิ่นอีสานใช้กฎทางวากยสัมพันธ์ (Syntax) หรือไวยากรณ์ (Grammar) ที่บอกความเกี่ยวข้องของคำในประโยคเหมือนกับภาษาไทยมาตรฐาน แต่มีไวยากรณ์ ได้แก่ ชนิดของคำ การสร้างคำ และโครงสร้างประโยค ที่มีเอกลักษณ์เฉพาะเป็นของตนเอง (อมรา ประสิทธิ์รัฐสินธุ์, 2544) ดังนี้

2.3.1.1 การออกเสียงเสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ต่างจากภาษาไทยมาตรฐาน และภาษาถิ่นอีสานไม่มีหน่วยเสียงควบกล้ำในขณะที่ในภาษาไทยมาตรฐานมี เช่น “ลูกกลอน” “กลางคืน” หน่วยเสียงในภาษาไทยถิ่นอีสาน คือ “ลูกกอน” “กางคืน” เป็นต้น

2.3.1.2 ระบบคำ ได้แก่ คำลงท้ายประโยคหรือวลี ภาษาไทยมาตรฐานใช้คำลงท้าย “สิ” “นะ” “คะ” “ครับ” ภาษาถิ่นอีสานใช้ “เด้อ” “นอ” “แน” “แม” เป็นต้น มีคำศัพท์ที่มีความหมายเดียวกันแต่พูดต่างกัน เช่น “พูด” ภาษาไทยมาตรฐาน ในภาษาถิ่นอีสานใช้ “เว้า” หรือ “ปาก” เป็นต้น มีคำที่เขียนเหมือนกันแต่ความหมายต่างกัน (พ้องรูป) เช่น “แพ้” ภาษาไทยมาตรฐาน ในภาษาถิ่นอีสานหมายถึง “ชนะ” เป็นต้น และระบบการสร้างคำ หรือไวยากรณ์ แตกต่างกัน เช่น การเรียงลำดับคำสลับกัน เช่น ภาษาไทยมาตรฐานใช้ “ส้มตำ” ภาษาถิ่นอีสานใช้ “ตำส้ม” เป็นต้น

2.3.2 ไวยากรณ์ภาษาไทยถิ่นอีสาน

ไวยากรณ์ภาษาไทยถิ่นอีสานเกี่ยวกับรูปคำและระเบียบในการประกอบรูปคำให้เป็นประโยคภาษาไทยถิ่นอีสาน (ชนานันท์ ตรงจิต, 2554; วิไลศักดิ์ กิ่งคำ, 2556) แบ่งเป็น 2 ระบบ ดังนี้

2.3.2.1 ระบบเสียงของภาษาถิ่นอีสาน

ประกอบด้วยหน่วยเสียง 3 ประเภท ได้แก่ เสียงพยัญชนะ สระ และวรรณยุกต์ หน่วยเสียงแต่ละประเภทมีลักษณะ ดังนี้

1) หน่วยเสียงพยัญชนะ แบ่งเป็น 2 ประเภท ได้แก่ หน่วยเสียงพยัญชนะเดี่ยว และหน่วยเสียงพยัญชนะควบกล้ำ ซึ่งพยางค์ (คำที่ประสมด้วยพยัญชนะ สระ วรรณยุกต์) ในภาษาถิ่นอีสานจะมีหน่วยเสียงเป็นเสียงพยัญชนะเดี่ยวเกือบทั้งหมด ส่วนหน่วยเสียงพยัญชนะควบกล้ำมีเฉพาะบางจังหวัด (สกลนคร นครพนม มุกดาหาร)

หน่วยเสียงพยัญชนะเดี่ยวยมีทั้งหมด 20 หน่วยเสียง แบ่งเป็นหน่วยเสียงพยัญชนะเดี่ยวที่เกิดต้นพยางค์ และเกิดได้เฉพาะท้ายพยางค์ (9 หน่วยเสียง) เกือบทั้งหมดเป็นเสียงพยัญชนะเหมือนกับภาษาไทยมาตรฐาน และหน่วยเสียงพยัญชนะภาษาไทยมาตรฐานที่ไม่มีในภาษาถิ่นอีสานมี 2 หน่วยเสียง ได้แก่ หน่วยเสียงพยัญชนะ ช (/ch/) และ หน่วยเสียงพยัญชนะ ร (/r/) โดยหน่วยเสียงพยัญชนะ ช (/ch/) ในภาษาถิ่นอีสานจะมีหน่วยเสียงเป็น ช ส ฉ (/s/) และ หน่วยเสียงพยัญชนะ ร (/r/) ในภาษาถิ่นอีสานจะเป็นหน่วยเสียงพยัญชนะ ฮ (/h/) และ ล /l/ และหน่วยเสียงพยัญชนะที่มีเฉพาะในภาษาถิ่นอีสาน ไม่มีในภาษาไทยมาตรฐานคือ หน่วยเสียง ย นาลิก

ตัวอย่างคำในภาษาไทยถิ่นอีสานมีหน่วยเสียงพยัญชนะเป็น ซ หรือ ส (/s/) เช่น

<u>ภาษาไทยมาตรฐาน</u>	<u>ภาษาไทยถิ่นอีสาน</u>
ซ้าง	ซ้าง
เจย	เสย
จีก	สีก/จิก

ตัวอย่างคำในภาษาไทยถิ่นอีสานมีหน่วยเสียงพยัญชนะเป็น ฮ (/h/) และ ล /l/ เช่น

<u>ภาษาไทยมาตรฐาน</u>	<u>ภาษาไทยถิ่นอีสาน</u>
ร้าย	ฮ้าย
เรื่อน	เฮื่อน
ราก	ฮาก
รถ	อ่านว่า “ลค” /lop ⁴ /
โรค	อ่านว่า “โลก” /lok ⁶ /

ตัวอย่างคำในภาษาไทยถิ่นอีสานที่มีหน่วยเสียงพยัญชนะ ย นาสิก เช่น

<u>ภาษาไทยมาตรฐาน</u>	<u>ภาษาไทยถิ่นอีสาน</u>
หญ้า	หญ้ำ (ออกเสียง ย นาสิก)
ยาก	ญาก (ออกเสียง ย นาสิก)
ใหญ่	ไหญ่ (ออกเสียง ย นาสิก)
อย่า	อย่า
อยู่	อยู่
อย่าง	อย่าง
อยาก	อยาก

ตัวอย่างการใช้พยัญชนะในภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐานที่มีหน่วยเสียงพยัญชนะในลักษณะเป็นปฏิภาคกัน (Sound Correspondence) เช่น

<u>ภาษาไทยมาตรฐาน</u>	<u>ภาษาไทยถิ่นอีสาน</u>
ผด	ฟด
คุ่ม	จุม
ดี (น้ำดี)	ปี
โนน	โพน

หน่วยเสียงพยัญชนะควบกล้ำ การใช้พยัญชนะควบกล้ำในภาษาไทยถิ่นอีสานแบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มพยัญชนะควบกล้ำที่ไม่พบการใช้เลยในทุกจังหวัด ได้แก่ กลกร คล คร ปล ปร ผล พล และ พร แต่จะใช้ในลักษณะลดเสียงพยัญชนะควบกล้ำ เช่น “ปลา” ลดเสียงเป็น “ป่า” “แพร” ลดเสียงเป็น “แพ” เป็นต้น และกลายเป็นคำอื่น เช่น “ปราบ” กลายเป็น “ผาบ” “กรอง” กลายเป็น “ทอง” “มะกรูด” กลายเป็น “หมากกูด” เป็นต้น และพยัญชนะควบกล้ำที่พบการใช้น้อย มีเฉพาะบางจังหวัด ได้แก่ กว คว ขว และ อว ในขณะที่จังหวัดอื่นที่ไม่ใช้พยัญชนะควบกล้ำกลุ่มนี้จะใช้ใน 3 ลักษณะ ได้แก่ ลดเสียงพยัญชนะควบกล้ำ เช่น “เกวียน” ลดเสียงเป็น “เกียน” “แก่ว” ลดเสียงเป็น “แก่ง” “แขวน” ลดเสียงเป็น “แขน” เป็นต้น และใช้เป็นสระเลื่อนสองเสียง ได้แก่ สระ เอีย เอือ อัว เช่น “กว่า” ใช้เป็น “กัว” “ขวาง” ใช้เป็น “ขวง” “ควม” ใช้เป็น “ควม” เป็นต้น และใช้หน่วยเสียงต่างไปจากเดิม เช่น “จวิด” ใช้เป็น “จิด” เป็นต้น

2) หน่วยเสียงสระ ภาษาไทยถิ่นอีสานมีหน่วยเสียงสระคล้ายคลึงกับภาษาไทยมาตรฐาน แบ่งเป็น 2 กลุ่ม ได้แก่ สระเดี่ยว และสระประสม หรือ สระเลื่อน โดยหน่วยเสียงสระเดี่ยวนั้นมี 18 หน่วยเสียง แบ่งเป็น 2 ประเภทย่อย ได้แก่ หน่วยเสียงสระเสียงสั้น 9 หน่วยเสียง ได้แก่ อะ อี อือ อุ เอะ แอะ เออะ โอะ เอาะ ตัวอย่างเช่น “อิม” “สุก” “เต็ม” “เติก” (กดเน้น) “แข็ง” (ตั้ง) “ขม” เป็นต้น และหน่วยเสียงสระยาว 9 หน่วยเสียง ได้แก่ อา อือ อู เอ แอ เออ โอ ออ ตัวอย่างเช่น “ฟ้า” “ลิง” (ร่างกาย) “แต่ง” “มินตา” (ลิมตา) “เลิก” (ถลอก) “สูบ” “โตน” (กระโดดลง) เป็นต้น และสระประสม หรือ สระเลื่อน มี 3 หน่วยเสียง ได้แก่ เอีย เอือ อัว ตัวอย่างเช่น “เดือน” “สวน” “เปือก” (เปลือก) “จัว” (วัว) เป็นต้น โดยจังหวัด มหาสารคาม ร้อยเอ็ด ไม่ปรากฏเสียงสระเอือ แต่จะใช้เสียงสระเอีย เช่น “เดือน” (เดือน) “เอียน” (เรือน) “เมี่ย” (เมือย) เป็นต้น

3) หน่วยเสียงวรรณยุกต์ ภาษาไทยถิ่นอีสานมีเสียงวรรณยุกต์ 5 - 6 หน่วยเสียง ขึ้นอยู่กับจังหวัด โดยจังหวัดที่มีหน่วยเสียงวรรณยุกต์ 5 หน่วยเสียง ได้แก่ หนองคาย สกลนคร นครพนม บึงกาฬ และมุกดาหาร จังหวัดที่มีหน่วยเสียงวรรณยุกต์ 6 หน่วยเสียง เช่น จังหวัดอุบลราชธานี (คำชัย ทองหล่อ, 2556) จากความแตกต่างของระดับเสียงวรรณยุกต์ในแต่ละจังหวัดส่งผลต่อการถอดเสียงวรรณยุกต์ในภาษาไทยถิ่นอีสานจำเป็นต้องมีการระบุว่าเป็นสำเนียงจังหวัดอะไร อย่างไรก็ดี การถอดเสียงและสะกดคำภาษาไทยถิ่นอีสานตรงตามสำเนียงพูดในแต่ละจังหวัดด้วยระบบการเขียนภาษาไทยมาตรฐานนั้น จะมีความยุ่งยากในเรื่องการสะกดคำ และยากต่อการอ่าน โดยรูปแบบการผันเสียงวรรณยุกต์และการสะกดคำที่ง่ายต่อการอ่านให้ตรงสำเนียงในแต่ละจังหวัด โดยเฉพาะคนอีสานนั้น คือการผันวรรณยุกต์ตามเกณฑ์การถอดเสียงและสะกดคำตามการเขียนภาษาไทยมาตรฐาน (ชนานันท์ ตรงดี, 2554)

2.3.2.2 โครงสร้างคำภาษาไทยถิ่นอีสาน

“คำ” ความหมายตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง เสียงพูด เสียงที่เปล่งออกมาครั้งหนึ่ง ๆ หรือหมายถึง เสียงพูด หรือลายลักษณ์อักษรที่เขียนหรือพิมพ์ขึ้นเพื่อแสดงความคิด มีความหมายในตัวและใช้ประกอบหน้าคำอื่น หรือ หมายถึงอักษรที่ประสมกันแล้วออกเสียงมาเป็นหน่วยเสียงเดียว หรือประสมด้วยสระตัวเดียว เรียกว่า พยางค์ โดยพยางค์ที่มีความหมาย เรียกว่า คำ หรือถ้อยคำ ถ้าเปล่งออกมาเป็นเสียงพูดเรียก คำพูด ถ้าเขียนเป็นตัวหนังสือ เรียก คำเขียน (กำชัย ทองหล่อ, 2543) และ คำภาษาไทยถิ่นอีสาน หมายถึง คำพูด หรืออักษรที่ประสมกันเป็นคำในแบบคำพยางค์เดียว หรือหลายพยางค์ เป็นภาษาไทยถิ่นอีสาน ที่มีความหมายอย่างใดอย่างหนึ่ง โดยลักษณะโครงสร้างคำภาษาไทยถิ่นอีสาน แบ่งประเภทตามจำนวนพยางค์ แบ่งได้เป็น 3 ประเภท (กำชัย ทองหล่อ, 2545) ดังนี้

- 1) โครงสร้างคำพยางค์เดียว เช่น “จ้าว” (ว้าว) “เฮด” (ท่า) “เบ็ง” (ดู) (ย่า) เป็นต้น
- 2) โครงสร้างคำสองพยางค์ เช่น “สะเอ้อ” (สะอึก) “กะทอ” (ภาชนะเครื่องจักสานชนิดหนึ่ง) “กะปอม” (กิ่งก่า) “กะบุง” (ตะกร้า) “กับกั๊ก” (ตุ๊กแก) เป็นต้น
- 3) โครงสร้างคำมากพยางค์ เช่น “กะคันขา” (ไช้คัน) “กะโบงตา” (กระบอกตา) “สามปี” (สารพัด) เป็นต้น

2.3.2.3 ชนิดคำภาษาไทยถิ่นอีสาน

ชนิดของคำในภาษาไทยถิ่นอีสานแบ่งตามทฤษฎีไวยากรณ์ศัพท์การก (Lexicase grammar) แบ่งเป็น 8 ประเภท (ชนานันท์ ตรงจิ, 2554) ดังนี้

- 1) คำนาม (Noun) คือ คำที่สามารถปรากฏหน้า หรือหลังคำกริยาได้ ปรากฏหลังคำบุพบทได้ หรือปรากฏหน้าคำคุณศัพท์ตัวกำหนด “นี้” “นั่น” ได้ และปรากฏหลังคำว่า “บ่” ไม่ได้ แบ่งเป็น 9 ชนิด ดังนี้

- 1.1) คำนามสามัญ (Common noun) หรือ สามานยนาม หมายถึง คำนามที่ใช้เรียกสิ่งต่าง ๆ ทั่วไป โดยไม่เจาะจงว่าเป็นสิ่งใดสิ่งหนึ่ง หรือคนใดคนหนึ่ง เช่น “อี่เก็ง” (ดวงเดือน) “จี่ฟ้า” (เมฆ) “หินแห่” (หินแสง) “ฟาน” (กวาง) “ดั่ง” (จมูก) เป็นต้น

- 1.2) คำนามชื่อเฉพาะ (Proper name) หรือ วิสามานยนาม หมายถึง คำนามที่เป็นชื่อเรียกเฉพาะของคน สัตว์ สถานที่ และสิ่งของ แบบเจาะจง เช่น “ป่าทอง” (ชื่อเฉพาะของคน) “บ้านโคกสี” (ชื่อเฉพาะของสถานที่) เป็นต้น

1.3) คำบุรุษสรรพนาม (Personal pronoun) หมายถึง คำนามที่ใช้เรียกแทนคำนามบุรุษที่ 1 - 3 ที่ปรากฏตามลำพัง และปรากฏเป็นส่วนหลักของนามวลีที่ประกอบด้วยคำปริมาณและคำลักษณนามได้ เช่น สรรพนามบุรุษที่ 1 เช่น “กู” “เฮา” “จ้อย” เป็นต้น สรรพนามบุรุษที่ 2 เช่น “มึง” “สู” “เจ้า” เป็นต้น และสรรพนามบุรุษที่ 3 เช่น “มัน” “เขา” “ลา/เพิ่น” เป็นต้น ซึ่งบุรุษสรรพนามเหล่านี้ สามารถเติมคำว่า “ชุม” “หมู่” “พวก” เข้าไปข้างหน้าเพื่อให้เป็นพหูพจน์ได้ และคำว่า “เจ้าของ” (ตัวเอง) ใช้เป็นสรรพนามบุรุษที่ 1 - 3 ได้

1.4) คำสรรพนามชี้เฉพาะ (Definite pronoun) หมายถึง สรรพนามที่ใช้แทนคำนามที่เป็นประธาน หรือ กรรมของกริยา ที่มีระยะห่างจากผู้พูด ได้แก่ ไก่ตัวผู้พูด เช่น “हीं” (นี่) “พี” (นี่) ห่างตัวผู้พูด เช่น “हीं” (นั่น) “พุ้น” (นั่น) ไก่ตัวผู้พูดและผู้ฟัง เช่น “พุ้น” เป็นต้น

1.5) คำสรรพนามถาม (Interrogative pronoun) ได้แก่คำว่า “อีหญ้ง” (อะไร) “ไผ” (ใคร) “ใส” (ไหน) เป็นต้น

1.6) คำสรรพนามไม่ชี้เฉพาะ (Indefinite pronoun) ได้แก่ “คำว้า” (อะไร) “ไผ” (ใคร) “ใส” (ไหน) มักปรากฏร่วมกับ “กะ” (ก็) ตัวอย่างเช่น “ไผกะได้ขอแต่ให้เป็นคนดี” (ใครก็ได้ขอให้เป็นคนดี) เป็นต้น

1.7) คำลักษณนาม (Classifier) หมายถึง คำนามที่ปรากฏร่วมกับคำจำนวน ทำหน้าที่เป็นส่วนขยายคำจำนวน หรือเป็นคำที่ปรากฏระหว่างคำนามกับคำคุณศัพท์ “นี่” “นั่น” ตัวอย่างเช่น “บักโมสองหน่วย” (แดงโม 2 ลูก) “บักโมนหน่วยนี้ใหญ่กว่า” (แดงโมลูกนี้โตกว่า) เป็นต้น

หากจำแนกคำลักษณนามตามคุณสมบัติทางอรรถศาสตร์ (ความหมาย) แบ่งเป็น 3 ประเภท ได้แก่ (1) ลักษณนามตามความหมาย ใช้กับมนุษย์และไม่ใช่มนุษย์ เช่น “คนสองคน” “ปลาสองโต” (ปลา 2 ตัว) เป็นต้น (2) ลักษณนามตามการเป็นกลุ่มและไม่เป็นกลุ่ม ใช้กับมนุษย์และไม่ใช่มนุษย์ เช่น “คนสองหมู่” (คน 2 กลุ่ม) “ปลาสองคอก” (ปลาสองครอก) เป็นต้น และลักษณนามแบ่งตามรูปทรง (กว้าง ยาว กลม ย่อย) พบเฉพาะที่ใช้กับสิ่งที่ไม่ใช่มนุษย์ เช่น “ผ้าสองผืน” “หมากพร้าวสองแขง” (มะพร้าว 2 ทะลาย) “ไผ่สองกอ” เป็นต้น (3) ลักษณนามตามที่มา ได้แก่ ลักษณนามที่ใช้ซ้ำกับคำนามที่ถูกขยาย เช่น “ปิ้งปลาสองปิ้ง” (ปิ้งปลาสองไม้) “กกไม้สองกก” (ต้นไม้สองต้น) เป็นต้น ลักษณนามที่กำหนดขึ้นจากคำเรียกอวัยวะ เช่น “ข่าสองหัว” “เหียนสองหลัง” (บ้านสองหลัง) เป็นต้น คำลักษณนามที่กำหนดขึ้นโดยเฉพาะคำนามอย่างหนึ่ง ๆ เช่น “แหสองตา” (แหสองผืน) “บ่วงสองคัน” (ช้อนสองคัน) เป็นต้น ลักษณนามที่มาจากการเล่นเสียง เช่น “แหงสองจิก” “ตดสองปูด” และลักษณนามที่มาจากคำกริยา เช่น “พีนสองมัด” “ยาสองหอ” เป็นต้น

นอกจากนี้ มีคำลักษณะนามที่เป็นคำลักษณะนามทั่วไป (General Classifier) สามารถใช้แทนคำนามทุกประเภท หรือใช้เมื่อไม่แน่ใจว่าจะใช้ลักษณะนามคำใด สำหรับอ้างอิงสิ่งที่กำลังกล่าวถึง คือ “อัน” เช่น “ข้าวจี๊ชาวอัน” (ข้าวจี๊ 20 อัน) ซึ่งปกติลักษณะนามของ ข้าวจี๊ คือ “ก้อน” แต่สามารถใช้ “อัน” แทนได้ เป็นต้น

1.8) คำนามแปลง (Nominalized noun) หรือ อาการนาม หมายถึง คำนามที่แปลงมาจากคำกริยาโดยการเติมอุปสรรค (Prefix) “การ” หรือ “ความ” ซึ่งในภาษาไทย ถิ่นอีสานพบการใช้บ่อย เช่น “ความเจ็บ ความไข้” (การเจ็บ การไข้) เป็นต้น

1.9) คำจำนวน (Numeral noun) หมายถึง คำนามสำหรับบอกจำนวน มีลักษณะเป็นตัวเลข เช่น 1 5 9 500 เป็นต้น หรือเป็นคำ เช่น “หนึ่ง” “เอ็ด” “ชาว” (ยี่สิบ) “ฮ้อย” (ร้อย) “พัน” เป็นต้น

2) คำกริยา (Verb) คือ คำที่เกิดหลังคำว่า “บ่” ได้ เช่น “กิน” “เล่น” (วิ่ง) “ดี” “ใหญ่” “น้อย” (เล็ก) “หลาย” “มัก” (ชอบ) “แซบ” (อร่อย) เป็นต้น โดยคำที่ปรากฏตามหลังคำกริยาได้ ได้แก่ คำนามที่ทำหน้าที่เป็นกรรมหรือส่วนเติมเต็ม กริยาที่ทำหน้าที่เป็นส่วนเติมเต็ม และบุพบทวลี (บุพบท + นาม หรือ บุพบท + กริยา) คำกริยาแบ่งตามการปรากฏร่วมกับคำที่ตามหลัง มี 10 ประเภท ดังนี้

2.1) กริยาลำพัง หรือ อกรรมกริยา หมายถึง กริยาที่อยู่ลำพังได้โดยไม่ต้องมีกรรม นามวลี หรือมีส่วนเติมเต็มตามหลัง เช่น “นอน” “เญียว” (เยียว) “เหงา” (หงอย) “อูก่อง” (กั๊กกั๊ก) “โห่ง” (พุ่ง) เป็นต้น และยังรวมถึง คำคุณศัพท์ หรือ คำวิเศษณ์ (กริยาคุณศัพท์) เช่น “ดี” “งาม” (สวย) “หลาย” (มาก) “ใหญ่” “กว้าง” (กว้าง) เป็นต้น

2.2) สกรรมกริยา หมายถึง คำกริยาที่มีคำนามที่ทำหน้าที่เป็นกรรมตามหลัง (คำนามที่เป็นผู้ถูกกระทำ) หรือมีอนุประโยคตามหลัง ตัวอย่างเช่น “บ้อย” (ดำ) “เปิด” (เบือ) “หญ่อง” (ขี้) “บาย” (สัมผัสด้วยมือ) “หัวซา” (ใส่ใจ) “ถั่ง” (กระแทก) เป็นต้น และรวมถึงคำสกรรมกริยาที่ซ้ำคำได้ เช่น “มักมาก” (ชอบชอบ) “ก้วนกวน” (รบก้วน รบกวน) เป็นต้น

2.3) กริยาเติมเต็ม หมายถึง คำกริยาที่ตามด้วยคำนามที่ไม่ใช่กรรม หรือเรียกคำนามนี้ว่าส่วนเติมเต็ม (Complement) กริยาเติมเต็มเป็นกริยาชนิดที่อยู่ลำพังไม่ได้ ตัวอย่างกริยาเติมเต็มเช่น “ม่ม” (รอดพ้น) “ลั้ง” (ลั่น) “คือ” (เหมือน) “เป็น” “แม่น” (ใช่) “หิว” “แง้น” (เขิน) “เข้า” “ออก” “จิ้น” “ลง” เป็นต้น

2.4) กริยาสถานที่ หมายถึง กริยาที่ตามด้วยคำนามบ่งบอกสถานที่ ซึ่งไม่ได้ทำหน้าที่เป็นกรรม และไม่ใช่อุปสรรคเติมเต็ม เช่น “ไป” “มา” “เมื่อ” (กลับไป) “อยู่” “จิ้น” “ลง” “เข้า” “ออก” เป็นต้น

2.5) กริยาตามด้วยกริยา หมายถึง กริยาที่ต้องตามด้วยกริยาเสมอ แบ่งเป็น 2 ประเภทคือ กริยาช่วย (Auxiliary verb) หมายถึง กริยาที่ต้องตามด้วยกริยาหลักของประโยค เช่น “ต้อง” “เป็นตา” (นำ) “เคย” “อยาก” “มัก” (ชอบ) และกริยาทั่วไปที่ตามด้วยกริยาทั่วไป ทำหน้าที่เติมเต็มหรือขยายให้สมบูรณ์ เช่น “จบ” (ลอบ) “แนม” (จับตาคู) “กาย” (ผ่าน) “คืน” (กลับ) เป็นต้น ซึ่งกริยาที่ตามหลังคำเหล่านี้ อาจจะมีหนึ่งคำหรือหลายคำในแบบกริยาเรียง (Serial Verbs) (คำกริยาเรียงต่อกัน โดยไม่มีคำอื่นคั่นกลาง)

2.6) กริยาแบบ “กริยา + ให้ + กริยาอื่น” มีความหมายเชิงบังคับ เช่น “ญู่จะให้...” (ยูแรงแยงให้...) “แอ่วให้...” (อ้วนให้...) “ตั้งให้...” “คองให้...” (รอกคอยให้...) “จ่มให้...” (บ่นให้...) “ญอมให้...” (ขอมให้...) “กุมให้...” (บังคับให้...) เป็นต้น

2.7) กริยาแบบ “กริยา + ว่า + กริยาอื่น” มีความหมายเชิงพูด อธิบาย อ้างอิง และคำที่อยู่ตามหลังบุพบท “ว่า” จะทำหน้าที่เป็นส่วนเติมเต็มให้กริยาที่มีความหมายสมบูรณ์ เช่น “จ่มว่า...” (บ่นว่า...) “ฮ่านว่า...” (กลัวว่า...) “คิดว่า...” (คิดว่า...) เป็นต้น

2.8) กริยาแบบ “กริยา + สิ + กริยาอื่น” หมายถึง คำกริยาที่ตามด้วยคำว่า “สิ” (จะ) และตามด้วยคำกริยาอื่น มีความหมายเกี่ยวกับความต้องการ ทำหน้าที่เติมเต็มให้ประโยคมีความหมายสมบูรณ์ เช่น “อยากสิ...” (อยากจะทำ...) “ฮนสิ...” (ร่นจะ...) “คิดสิ...” (คิดจะทำ...) “เบ็งซงสิ...” (จะทำจะ...) เป็นต้น

2.9) กริยาแบบ “กริยา + บุพบท + คำนาม/นามวลี” หมายถึง อกรรมกริยาที่ต้องมีบุพบทเฉพาะและตามด้วยนามหรือนามวลี (กริยาที่เลือกบุพบท) เช่น “บรรลุด...” (บรรลุด ถึง...) “ตอบสนอง กับ...” (ตอบสนองต่อ...) “มา แต่...” (มาจาก...) เป็นต้น โดยคำกริยาเหล่านี้บางครั้งจะไม่ใช่ หรือละบุพบทเฉพาะ

2.10) กริยาแบบ “กริยา + กัน/กันกับ + นามวลี” ที่แสดงความหมายแฝงเกี่ยวกับการกระทำซึ่งกันและกันระหว่างคนตั้งแต่ 2 คนขึ้นไป เช่น “พ้อ” (พบ) “ญาด” (แย่ง) “ปะ” (หย่า) “ผิด” (ทะเลาะ) “คือ” (เหมือน) “ต่อ” “คุย” “จอด” (เชื่อม)

3) คำคุณศัพท์ (Adjective) หมายถึง คำที่ปรากฏหลังคำนามและ ปรากฏหลัง “บ่” “ไม่ได้” ทำหน้าที่ขยายคำนามนั้น แบ่งเป็น 2 ประเภท ดังนี้

3.1) คำคุณศัพท์ประเภทบอกกำหนด (Determining adjective) แบ่งตามลักษณะทางความหมาย 5 ประเภท ได้แก่ (1) คำคุณศัพท์ประเภทบอกลำดับ เช่น “คำอิด” (แรก) “ลุน” (หลัง) “หล้า” (สุดท้าย) “ก่อน” เป็นต้น (2) คำคุณศัพท์บอกปริมาณ เช่น “ทั้งเหมิด” (ทั้งหมด) “หนึ่ง” “เดียว” (3) คำคุณศัพท์บ่งชี้ เช่น “นี้” “ใด” “นั่น” เป็นต้น (4) คำคุณศัพท์บอกจำกัด เช่น “ท่อนั้น” เป็นต้น และ (5) คำคุณศัพท์บอกเน้น เช่น “เอง” เป็นต้น

3.2) คำคุณศัพท์บอกคุณลักษณะ (Qualitative Adjective) คือ คำคุณศัพท์ที่ปรากฏหลังคำนาม และเกิดจากการซ้ำคำกริยาคุณศัพท์ (Adjective Verb) เช่น “ขาว ๆ” “แดง ๆ” “ดำ ๆ” “หวาน ๆ” “เผ็ด ๆ” “อ้วน ๆ” “จ้อย ๆ” (ผอม ๆ) เป็นต้น

4) คำวิเศษณ์ (Adjective) คือ คำที่ปรากฏหน้าหรือหลังคำกริยา คำคุณศัพท์ คำวิเศษณ์ หรือหน้าประโยค คำวิเศษณ์ปรากฏคำว่า “บ” นำหน้าไม่ได้ ทำหน้าที่เป็นส่วนขยาย แบ่งได้ 4 ชนิด ดังนี้

4.1) คำวิเศษณ์หน้ากริยา ทำหน้าที่ขยายกริยาที่อยู่ข้างหลัง คำวิเศษณ์นั้น เช่น “ค่อย ๆ” “ลี” (จะ) “บ” (ไม่) “คือลี” (คงจะ) “เกือบ” “ลี” (ค่อนข้าง) เป็นต้น

4.2) คำวิเศษณ์หลังกริยา ทำหน้าที่ขยายกริยาที่อยู่ข้างหน้า คำวิเศษณ์นั้น แบ่งเป็น 2 ประเภทตามลักษณะการใช้เสียงสื่อความ พรรณภาพ หรือเหตุการณ์ซึ่งรับรู้ผ่านทางประสาทสัมผัส ตา ลิ้น จมูก หู ผิวหนังและร่างกาย หรือเรียกว่า คำอیدیโอโฟน (Ideophone) และคำวิเศษณ์หลังกริยาที่ไม่เป็นอیدیโอโฟน

คำวิเศษณ์หลังกริยาประเภทคำอیدیโอโฟน ได้แก่ พรรณาสงสลี เช่น “เลือดแดงจิ้งจิ้ง” (เลือดแดงฉาน) “แดงโมแดงจ่าย่วย” (แดงโมแดงฉ่ำ) พรรณารูปลักษณะ เช่น “โตเตี้ยต้อกม้อก” (ตัวเตี้ยม้กก) “ขยวออดหลาด” (ขยวระหง) พรรณาท่าทาง เช่น “นอนหมอบเอคเอค” (นอนคว่ำแผ่) “นอนขยชีเก้เก้” (นอนขยชยวขึ้นม) พรรณารส เช่น “ขมอ้ออ้อ” (ขมกลมกล่อม) “จางเจ่วเลว” (จืดกร่อย ๆ) พรรณากลิ่น เช่น “เหม็นตึ้ง” (เหม็นฉิ่ง) “หอมฮ่วยฮ่วย” (หอมกรุ่น) พรรณเสียง เช่น “เชือกขาดดั่งปุดบาดหนึ่ง” (เชือกขาดดั่งปุด) “เด็กน้อยโดคน้ำตุ้มตุ้ม” (เด็กน้อยกระโดคน้ำตุ้มตาม) พรรณาสัมผัสที่เป็นความรู้สึก เช่น “คิงฮ้อนฮุ่มฮุ่ม” (ตัวร้อนรุ่มรุ่ม) “มือเย็นจ้อยจ้อย” (มือเย็นเจี๊ยบ) และ พรรณาสัมผัสของสภาพวัตถุ เช่น “มือชยาคากคาก” (มือหยาบกร้าน) “ผ้าเป็ยกแมะแหละ” (ผ้าเป็ยกชื้น)

คำวิเศษณ์หลังกริยาประเภทไม่ใช่คำอیدیโอโฟน เช่น “ลาวเมือขยมบ้านดูชะนาค” (เขากลับไปเยี่ยมบ้านบ่อยมาก) “อย่าลึมไปน่านมกนเข้านาค” (อย่าลึมไปตามมันมกนเข้าด้วยนะ) เป็นต้น

4.3) คำวิเศษณ์ที่ปรากฏหน้าประโยค ทำหน้าที่ขยายประโยคเพื่อเชื่อมโยงความสัมพันธ์ เช่น “จั่งใดกะตาม” (อย่างไรก็ตาม) “จั่งใดกะดี” (อย่างไรก็ดี) เป็นต้น

4.4) คำวิเศษณ์หรือคำที่ทำหน้าที่เหมือนคำวิเศษณ์ แบ่งประเภทตามความหมาย แบ่งได้ 6 ประเภท ได้แก่ บอกอากัปกริยา เช่น “ไว ๆ” “เคี่ยม” (สงบเสงี่ยม) “ค่อย ๆ” (เบา ๆ / ค่อย ๆ) เป็นต้น บอกความถี่และซ้ำคำ เช่น “อยู่เลื่อย ๆ” (อยู่เรื่อย ๆ) “คู้คู้” (บ่อยบ่อย) “คี้คี้” (ดีมาก/บ่อย ๆ) เป็นต้น และคำวิเศษณ์วลี เช่น “บางเทื่อ” (บางครั้ง) “ทำมะดา” (ธรรมดา/โดยปกติ) เป็นต้น คำวิเศษณ์บอกเวลา เช่น “มื่อนี้” (วันนี้) “มื่ออื่น” (พรุ่งนี้) “มื่อตึ้ง” (มะเรื่องนี้) เป็นต้น คำวิเศษณ์บอกสถานที่ เช่น “อยู่ข้าง” เป็นต้น วิเศษณ์เน้นเสริมความ (Intensifier) ของกริยา ปรากฏ

หลังกริยาของประโยค เช่น “โพล” (เหลือเกิน) “กะดื้อ” (ขนาดนี้) “กะดัดกะดื้อ” (ขนาดนี้) “บักคัก” (อย่างยิ่ง) เป็นต้น และคำวิเศษณ์ประเภทอื่น ๆ เช่น “เอง” “นำ” (ด้วย) “แหง” (ข้าง) “คือกัน” (เหมือนกัน) เป็นต้น

5) คำบุพบท (Preposition) คือ คำที่ปรากฏหน้าคำนามในหน่วยภาษาที่ส่วนประกอบมีความสำคัญเท่ากัน (Exocentric construction) เป็นคำที่ไม่สามารถตามด้วยคำว่า “ใด” หรือ “อื่น” ได้ ไม่สามารถปรากฏอยู่ท้ายประโยคเน้นความ ไม่สามารถปรากฏร่วมกับคำปริมาณ และไม่สามารถปรากฏร่วมกับคำวิเศษณ์ปฏิเสธ “บ่” ได้ เช่น “เทิง” (บน/เหนือ) “ใต้” (ล่าง/ใต้) “กาง” (กลาง) “นอก” แต่ (จาก) และคำบุพบทที่กลายมาจากคำชนิดอื่น เช่น กลายมาจากคำนาม ได้แก่ “หน้า” “หลัง” “ข้าง” และกลายมาจากคำกริยา เช่น “นำ” (ตาม) เป็นต้น

6) คำสันธาน (Conjunction) คือ คำที่ปรากฏระหว่างคำหรือข้อความที่มีน้ำหนักเท่ากันให้ติดต่อกันเป็นเรื่องเดียวกัน และไม่ปรากฏหลังคำว่า “บ่” แบ่งเป็น 4 ประเภท ได้แก่

6.1) สันธานที่เชื่อมความคล้ายตามกัน ได้แก่ “กับ” (กับ/และ) “แล้วกะ” (แล้วก็) ตัวอย่างเช่น “กับเข้ามือนี้มีดัมกับแจ่ว” (กับข้าววันนี้มีดัมและน้ำพริก) “มันว่าบ่มาแล้วกะบ่คิดสิมาซ้า” (มันว่าไม่มา แล้วก็ไม่คิดจะมาด้วยซ้ำ) เป็นต้น

6.2) สันธานที่เชื่อมความแย้งกัน เช่น “แต่ว่า” “แต่ว่า...สิ้น” (แต่ว่า... กลับ) ตัวอย่างเช่น “บ่อยากไป แต่ว่ามันสั้นบ่บอกไผ” (ไม่อยากไปแต่ว่ามันกลับไม่บอกใคร) เป็นต้น

6.3) สันธานที่เชื่อมความให้ทางเลือก เช่น “หรือ” (หรือ) “หรือบ่...กะ” (หรือไม่...ก็) ตัวอย่างเช่น “เจ้าสิไปหรือบ่ไปเกาะ” (คุณจะไปหรือไม่ไปนะ) “มันสิไปอยู่หรือบ่กะบ่จัก” (มันจะอยู่หรือไม่ก็ไม่รู้) เป็นต้น

6.4) สันธานที่เป็นการบ่งชี้ ได้แก่ “คือ” “กะคือ” (ก็คือ) ตัวอย่างเช่น “เทิงสองโตคือจอนฟอนกับงูเห่า หาเลื่องกั๊กกันอยู่คู่มื้อ” (ทั้งสองคือฟองพอนกับงูเห่า หาเรื่องกั๊กกันอยู่ทุกวัน) “เลื่องต่อไปกะคือเลื่องทุละบ่แม่น” (เรื่องต่อไปก็คือเรื่องทุละไม่ใช่)

7) คำปริมาณ (Conjunction) คือ คำที่ปรากฏหน้าคำนามในหน่วยสร้างเข้าสู่ศูนย์ และไม่ปรากฏร่วมกับ บ่ เช่น หนึ่ง สอง สาม ลาง (บาง) แต่ละ หลาย เค็ง (ครั้ง) ตัวอย่างเช่น ลางเพื่อฮ้อนลางเพื่อเย็น (บางครั้งร้อนบางครั้งเย็น)

8) คำอนุภาค (Particle) คือ คำที่ปรากฏในตำแหน่งหน้าสุด เรียกว่า คำอุทาน และปรากฏหลังสุดของประโยค หรือถ้อยความ (Utterance) เรียกว่า คำลงท้าย คำอนุภาคจะไม่สัมพันธ์กับคำใดคำหนึ่งในประโยคเป็นการเฉพาะ และปรากฏหลังคำว่า “บ่” ไม่ได้ โดยคำอนุภาคในภาษาไทยถิ่นอีสานแบ่งเป็น 3 ประเภทดังนี้

8.1) คำอนุภาคต้นประโยค เช่น “อ๋น” (เอ่อ) “ฮ่วย” (เอะ) “เอา/เอ้า” (อ้าว) “ป้าด” (ไอ้โฮ) “ป้าดโท” (อูวะ) “บุญ” (อุ๊ย) “อื้อ” (เออ) “โธ” (อ้อ) “หึ” (ไม่) เป็นต้น

8.2) คำอนุภาคท้ายประโยค ได้แก่ ลงท้ายแสดงการถาม เช่น “บ่” (ไหม) “ติ” (หรือ) “ไป” (หรือยัง) เป็นต้น คำลงท้ายแสดงการสั่ง เช่น “โลด” (เลย) “เถาะ” (เถอะ) “แห่น” (หน่อย) คำลงท้ายแสดงการแจ้งให้ทราบ เช่น “เด้อ” “คอก” (หรือ) “ตั้ว” (นะ) เป็นต้น คำลงท้ายแสดงความรู้สึก เช่น ขะน้อย (ข้าน้อย) “จ๊ะ” “จ๋า” เป็นต้น นอกจากนี้ยังมีคำอนุภาคท้ายประโยคที่คำแรกแสดงเจตนาและให้ความหมายหลัก คำที่ 2 เป็นตัวเสริมความ คำที่ 1 และ คำที่ 3 เสริมความ คำที่ 1 และ 2 คำอนุภาคบางคำจะไม่ปรากฏเป็นอนุภาคเดี่ยว

ตัวอย่างคำอนุภาคที่เกิดร่วมกันตั้งแต่ 2 คำ แบบ “ถาม + ถาม” เช่น “ติบ” (มั้งนะ) “บ่หน่อ” (ไหมนะ) แบบ “สั่ง + สั่ง” เช่น “ชะเถาะ” (ชะเถอะ) “แน่เถาะ” (หน่อยเถอะ) แบบ “แจ้ง + แจ้ง” เช่น “คอกเด้อ” (คอกนะ) “ซ้าคะ” (ด้วยซ้า) “ซ้าเนา” (ด้วยซ้านะ) เป็นต้น แบบ “สั่ง + ถาม” เช่น “ชะติ” (ชะสิ) “ชะบ่” (ชะดีมั้ง) แบบ “สั่ง + แจ้ง” เช่น “แห่นเด้อ” (หน่อยนะ) “โลดเด้อ” (เลยนะ) แบบ “ถาม + ถาม + แจ้ง” เช่น “บ่บูเหลือว” (หรือเปล่าก็ไม่รู้) แบบ “สั่ง + สั่ง + แจ้ง” เช่น “เลยโลดเหลือว” (เลยนะซี) แบบ “สั่ง + สั่ง + ถาม” เช่น “ชะเถาะเนา” (ชะเถิดนะ) “แห่นแม่ละ” (ด้วยหน่อยซี) เป็นต้น แบบ “สั่ง + ถาม + ถาม” เช่น “แห่นติละ” (หน่อยไหม) “ชะติละ” (ชะซินะ) และแบบ “สั่ง + ถาม + แจ้ง” เช่น “แน่ติเว้ย” (หน่อยซีเว้ย) เป็นต้น

8.3) คำอนุภาคประเภทอื่น ได้แก่ คำอุทานประเภทแสดงเจตนาอื่น ๆ นอกเหนือการสั่ง แจ้ง ได้แก่ การตัดพ้อ เช่น “กะดาบ” (นี่ก็) “เหม” (ก็) ใช้เรียกขานบุคคล เช่น “เอ้ย” (เอ๊ย) ใช้ขานตอบ เช่น “เอ้ย” “หือ/เหอ” (อือ) เป็นต้น

2.3.3 การประสมคำภาษาไทยถิ่นอีสาน

การประสมคำเป็นวิธีการเพิ่มจำนวนคำในภาษาให้มีมากขึ้น โดยการนำคำต่าง ๆ ตั้งแต่ 2 คำ/หน่วยคำขึ้นไปมาประกอบเข้าด้วยกัน ด้วยวิธีนำคำมาเรียงต่อกัน ประสมกัน สมาสกัน ซ้อนคำ หรือ ซ้ำคำ โดยคำที่สร้างขึ้นมานี้จะเรียกชื่อตามวิธีการสร้างคำ ได้แก่ คำประสม คำสมาส และ คำซ้อน

2.3.3.1 คำประสม

คำประสม หมายถึง คำที่เกิดจากการนำคำตั้งแต่ 2 คำหรือหน่วยคำ มาเรียงต่อกันเป็นคำใหม่ที่มีความหมายต่างจากความหมายรวมของคำหรือหน่วยคำเดิม หรือยังมีเค้าความหมายเดิมอยู่ โดยลักษณะคำประสมในภาษาไทยถิ่นอีสาน แบ่งตามลักษณะทางความหมาย ได้ 3 ลักษณะ ได้แก่ แบบคงเค้าความหมายเดิม หรือเชิงเปรียบเทียบ เช่น “เข้าจี” (ข้าว + จี = ข้าวที่จีโรยเกลือ) “ก้องแขน” (กำไล + แขน = กำไลที่ใช้ใส่แขน) เช่น “เสื่อร้องไห้” (เนื้อติดมันย่าง)

“แจ่วฮ้อน” (สุกี้แบบอีสาน) เป็นต้น แบบคำจ่ากลุ่ม-ลูกกลุ่ม หรือการบอกส่วนหลักและส่วนย่อย เช่น “เข้า” (ข้าว) เป็นจ่ากลุ่ม และมีประเภทข้าวเป็นลูกกลุ่ม ดังนี้ “เข้าเหนียว” (ข้าวเหนียว) “เข้าก่ำ” (ข้าวเหนียวดำ) “เข้าจ้าว” (ข้าวเจ้า) เป็นต้น แบบกริยากลื่นความ หรือคำกริยาที่รวมความหมายของคำนามที่เป็นประธานหรือกรรมให้กลายเป็นคำใหม่ เสมือนเป็นคำเดียว เรียกว่า คำกริยากลื่นความ ประธาน เช่น “ปากกึก” (เป็นใบ้) “คังหวึก” (เพดานโหว่) เป็นต้น หรือ กริยากลื่นความกรรม เช่น “กั๊กมือ” (กั๊กมือ) “ส่วยหน้า” เป็นต้น

ชนิดของคำประสมในภาษาไทยถิ่นอีสาน แบ่งเป็น 2 ชนิด คือ

1) คำประสมชนิดคำกริยา ได้แก่ คำประสมที่มีโครงสร้างการประสมคำแบบ “คำกริยา+คำกริยา” เช่น “เหงานอน” (ง่วงสัปหงก) “จอบหลอย” (จ้องจขโมย) “ตายคีน” (พีนจากสลบ) เป็นต้น แบบ “คำกริยา+คำนาม” เช่น “ดับแนว” (สูญพันธ์) “เว้าพีน” (นินทา) “กันแจ้ว” (หมั่นเจี้ยว) เป็นต้น แบบ “คำกริยา+คำนาม+คำกริยา” เช่น “หลับตาซัง” (ค้อนปะหลับปะเหลือก) “ตั้งตาน้อย” (มองค้วยหางตา) เป็นต้น แบบ “คำกริยา+คำบุพบท” เช่น “บ้านนอก” (เขย) “เป็นทาง” (ไม่เข้าข้างฝ่ายใด) เป็นต้น แบบ “คำกริยา+คำบุพบท+คำนาม” เช่น “กินตามน้ำม” (กินตามน้ำ) เป็นต้น แบบ “คำนาม+คำกริยา” เช่น “จี้ลี้” (จี้เหนียว) “จี้ฮ้าย” (จี้เหว) เป็นต้น และแบบ “คำนาม+คำกริยา” เช่น “เลือดเข้าตา” (สู้อย่างจนมุม) “น้ำมถ่วมปาก” (พูดไม่ออกเพราะเกรงจะมีภัย) เป็นต้น

2) คำประสมชนิดคำนาม ได้แก่ คำประสมแบบ “คำนาม+คำนาม” เช่น “ไข่เข้า” (ไข่ข้าว) “บังไฟ” “จี้เพี้ย” (จี้อ่อนในไส้เล็กของวัว) แบบ “คำนาม+คำนาม+คำนาม” เช่น “ไบยานาง” “หลอดไฟฟ้า” แบบ “คำนาม+คำลักษณะนาม” เช่น “หนามแท่ง” “ซันพุด” (เนื้อวัวพร้อมเครื่องในที่แบ่งไว้เป็นชุด ๆ) “น้ำมขวด” (น้ำขวด) แบบ “คำนาม+คำกริยา” เช่น “พ้อฮ้าง” (พ้อหม้าย) “ไล่ตัน” แบบ “คำนาม+คำกริยา+คำนาม” เช่น “ไข่ฮ้างฮ้าง” (ไข่ที่ไม่ฟักเป็นตัว) “ผ้าเคียนแอว” (ผ้าผูกแอว) “ค้อนตอนตะ” ปู (ค้อน) แบบ “คำนาม+คำกริยา+คำกริยา” เช่น “น้ำมแข็งใส” (น้ำแข็งใส) แบบ “คำนาม+คำบุพบท+คำนาม” เช่น “ลูกในบ้าน” (ลูกหลานในหมู่บ้าน) “บัวได้น้ำม” (คนไม่รู้ธรรมะ) แบบ “คำกริยา+คำกริยา” เช่น “ซอยแซ่” (ลาบดิบ) “ตำส้ม” (ส้มตำ) “ขอทาน” แบบ “คำกริยา+คำนาม” เช่น “ปิ้งไก่” (ไก่อ่าง) “ส้มผัก” (ผักดอง) แบบ “คำนาม+คำบุพบท” เช่น “เสื่อใน” (เสื่อชั้นใน) “เครื่องใน” (อวัยวะภายใน) แบบ “คำบุพบท+คำนาม” เช่น “ใต้ล่าง” (ข้างล่าง) “เตียงเฮือน” (ชั้นบนของบ้าน) แบบ “คำนาม+คำวิเศษณ์เลียนเสียง” เช่น “บั้งกะโปก” (ประทัด) “เกิบเตะ” (รองเท้าเตะ) แบบ “คำนาม+จำนวน” เช่น “แจวซาว” (พื้นที่ที่ยี่สิบ) “ขันแปด” (ขันใส่เครื่องไหว้แปดอย่าง) และแบบ “จำนวน+คำนาม” เช่น “สามสิบทิบ” (ผ้าขี้ริ้ว) “สามง่าม” (ไม้สามง่าม) เป็นต้น

2.3.3.2 คำสมาส

คำสมาส หมายถึง คำที่สร้างโดยนำคำมาเชื่อมต่อกัน คำหลักจะอยู่หลัง คำขยาย มักออกเสียงประวิตรชนิษฐ์ระหว่างคำที่นำมาประกอบกัน คำสมาสในภาษาไทยอีสาน ส่วนใหญ่จะยืมมาจากภาษาไทยมาตรฐาน เช่น พุทธศาสนา (พุทธ + ศาสนา) รัฐบาล (รัฐ + บาล) ราชการ (ราช + การ) เป็นต้น

2.3.3.3 คำซ้อน

คำซ้อน คือ การนำคำตั้งแต่ 2 คำที่เหมือนกัน คล้ายกัน ทำนองเดียวกัน ตรงข้ามกัน มาเรียงต่อกันเพื่อสร้างคำใหม่ มักซ้อนคำเป็นคู่ ๆ แบบ 2 คำ 4 คำ หรือ 6 คำ โดยแต่ละคำจะมีน้ำหนักทางความหมายเท่ากัน อาจมีสัมผัสและไม่มีสัมผัสกันก็ได้ และการสัมผัสกันอาจเป็นการสัมผัสกันระหว่างส่วนหน้ากับส่วนหลัง สัมผัสกันในแบบคู่ขนาน หรือกล่าวซ้ำกับคำที่ 1 กับคำที่ 3 รวมถึงการนำคำในภาษาต่าง ๆ มาซ้อนกันเพื่อสร้างคำ ตัวอย่างคำซ้อนในภาษาไทยถิ่นอีสาน แบบต่าง ๆ มีดังนี้

คำซ้อนที่มีความหมายเหมือนกัน เช่น มิตรสหาย เฟ่งเล้ง เป็นต้น

คำซ้อนที่มีความหมายในทำนองเดียวกัน เช่น ข่มเหง ชักชวน (ชักชวน) ไส้หนา (ไธ่หนา) เป็นต้น

คำซ้อนที่มีความหมายตรงกันข้าม เช่น ผิดถึก (ผิดถูก) พอดีพอฮ้าย (พอดีพอร้าย) ฮ่วมทุกข์ฮ่วมสุข (ร่วมทุกข์ร่วมสุข) เป็นต้น

คำซ้อนที่สัมผัสกันแบบ 2 คำ และ 4 คำ เช่น “คั่นคาย” (ระคายคั่น) “ฟ้าวฝั่ง” (เว้งริบ) “อกใหม่ไส้ขม” “จีตักมกได้” (ชอบตักดวง) เป็นต้น

คำซ้อนที่ไม่มีสัมผัสแบบ 2 คำ และ 4 คำ เช่น “ซ้างม้า” (ซ้างม้า) “เก็บเมี้ยน” (เก็บรักษา) “ฮักแพง” (รักใคร่ถนุถนอ) “กุ่งหอยปูปลา” (กุ่งหอยปูปลา) “อืดอยากทุกญา” (อดอยากทุกข์ยาก) เป็นต้น

คำซ้อนที่มีความหมายคู่ขนานแบบ 4 คำ เช่น “ตั้นเด็กลูกเช่า” (ตั้นเด็กลูกเช่า) “บ้านเกิดเมืองนอน” เป็นต้น

คำซ้อนแบบกล่าวซ้ำคำที่ 1 กับคำที่ 3 เช่น “กางเก้ากางสิบ” (ครึ่ง ๆ กลาง ๆ) “กินฮื้อกินเฮง” (เอารัดเอาเปรียบ) “ตาซู้ตาซัน” (จ้องมองจนตาไม่กระพริบ) เป็นต้น

คำซ้อน 6 คำ โดยคำบางคำอาจมีมากกว่า 1 พยางค์ แบ่งเป็น 2 ส่วน ส่วนละ 3 คำ มักมีเสียงสัมผัสกันระหว่างส่วนหน้ากับส่วนหลัง และกล่าวซ้ำกันอยู่ในทั้งสองส่วน เช่น “เลือกหม้องฮัก มกหม้องซัง” (เลือกที่รักมกที่ซัง) คำกล่าวซ้ำคือ “หม้อง” และคำที่สัมผัสกันคือ “ฮัก” สัมผัสกับ “มก” หรือ “มอบไส้ตัก หักไส้พา” (ยกให้เป็นกรรมสิทธิ์) คำกล่าวซ้ำคือ “ไส้” และคำที่สัมผัสกันคือ “ตัก” สัมผัสกับ “หัก” เป็นต้น

คำซ้อนจากการนำคำในภาษาต่าง ๆ มาซ้อนกัน เช่น คำภาษาไทยถิ่นอีสานซ้อนคำภาษาไทยมาตรฐาน (คำที่ขีดเส้นใต้เป็นคำภาษาไทยถิ่นอีสาน) เช่น “แก่เ้า” (แก่เ้า) “เก็บำ” “พดวี” เป็นต้น คำภาษาไทยมาตรฐานซ้อนกับคำภาษาไทยมาตรฐาน เช่น “ฝนฟ้า” “กดจี่” “จ็องแวะ” “ญกเว้น” (ยกเว้น) เป็นต้น คำภาษาไทยซ้อนกับภาษาบาลี และภาษาเขมร (คำที่ขีดเส้นใต้คือคำภาษาไทย) เช่น “ข้าทาด” (ข้าทาด) “ทุกญาก” (ทุกขยาก) เป็นต้น คำบาลีซ้อนกัน คำเขมรซ้อนกัน หรือคำบาลีซ้อนคำเขมร เช่น “บุนคุณ” (บุญคุณ) “สีนทำ” (สีลธรรม) เป็นต้น

ชนิดของคำซ้อน แบ่งได้เป็น 2 ชนิด ได้แก่

- 1) คำซ้อนชนิดคำนาม เช่น “เขตแน” “ลีววน” (ลีววาน) “กุองกุมี่” (กุมงกุมี่) “ถ้วยซาม” (ถ้วยซาม) “เด็กเล็กเด็กน้อย” (พวกเด็ก) มิตรสหาย เป็นต้น
- 2) คำซ้อนชนิดคำกริยา เช่น “ข็องเกียว” (เกิยวข็อง) “กิดฮอดกิดถิง” (กิดถิง) “แงซักหักกะดูก” (แงซักหักกะดูก) “ปำฮ็อง” (ปำวร็อง) “เบ็งแญง” (ดูแล) เป็นต้น

2.3.5.4 คำซ้ำ

คำซ้ำ หมายถึง การสร้างคำโดยนำคำเดิมมาซ้ำกันตั้งแต่ 2 ครั้ง ทำให้ได้คำใหม่ที่อาจมีความหมาย เปลี่ยนน้ำหนัก หรือชนิดของคำ ต่างไปจากคำเดิม หรือคงชนิดของคำเดิมไว้ แบ่งเป็น 2 ประเภท ดังนี้

- 1) คำซ้ำแบบคงชนิดคำแบบเดิม ได้แก่ คำซ้ำชนิดคำนามบอกความเป็นพหูพจน์ เช่น “ดูก ก ก ลาน ก” “อ้าย ก น็อง ก” (พี ก น็อง ก) เป็นต้น ชนิดคำนามบอกการนับแยกเป็นส่วน ๆ เช่น “กะสอบ ก” (กะสอบ ก) “ต่อน ก” (ชีน ก) เป็นต้น และคำซ้ำชนิดคำนามบอกการประมาณการจำนวน เช่น “พัน ก” “หมีน ก” เป็นต้น คำซ้ำชนิดคำกริยาบอกความหมายการกระทำซ้ำ ๆ อย่างต่อเนื่อง เช่น “พัง ก” “นั่ง ก” “นอน ก” เป็นต้น และคำซ้ำชนิดคำกริยาบอกการเร่งกระทำให้เสร็จโดยเร็ว เช่น “กิน ก” “เกา ก” เป็นต้น คำซ้ำชนิดคำคุณศัพท์ใช้เน้นลำดับที่ หรือปริมาณ เช่น “ตำอิด ก” (รแก ก) “กั่ว ก” (กวา ก) เป็นต้น คำซ้ำชนิดคำคุณศัพท์ใช้เน้นปริมาณ เช่น “เกือบ ก” “หลาย ก” เป็นต้น คำซ้ำชนิดคำวิเศษณ์ ใช้เน้นความหมาย เช่น “ค่อย ก” “จัง ก” “แ้ ก” และคำซ้ำชนิดคำวิเศษณ์ใช้บอกการกระทำต่อเนื่อง เช่น “เลื่อย ก” (ร็อย ก) “ตั้ง ก” (ถิง ก) “จ็อก ก” (หงึก ก) เป็นต้น

- 2) คำซ้ำแบบเปลี่ยนชนิดคำ เช่น

คำกริยาที่มาจากกรซ้ำคำนาม เช่น “เลื่องหมก” (ร็องหม ก) เป็นต้น

คำกริยาที่มาจากกรซ้ำคำคุณศัพท์ เช่น “ตำหมากหุงแบบ”

(สัมคำอรอยอรอย) เป็นต้น

คำคุณศัพท์ที่มาจากกริยาคำกริยา บอกความไม่ชัดเจน เช่น “อันน้อย ๆ” (อันเล็ก ๆ) “หน้าแหล่ ๆ” (หน้าคล้ำ ๆ) เป็นต้น ไม่ระบุระยะเวลาชัดเจน เช่น “งามเด็ก ๆ” (ตอนเด็ก ๆ) “งามแดง ๆ” (เวลาเย็น ๆ) หรือบอกเน้นความหมายคำกริยาเดิม เช่น “หน่วยหิม ๆ” (ลูกห่าม ๆ) “ปาเป็น ๆ” (ปลาเป็น ๆ) เป็นต้น

คำคุณศัพท์ที่มาจากกริยาคำนาม ให้ความหมายใหม่ เช่น “ของก้วย ๆ” (ของง่าย ๆ) หรือให้ความหมายโดยประมาณ เช่น “เนวกินน้ำ ๆ” (อาหารน้ำ ๆ) “หน้าตาแขก ๆ” เป็นต้น คำบุพบทที่มาจากกริยาคำนาม ให้ความหมายแบบไม่เจาะจง เช่น “อยู่แถว ๆ เอือนนี้” (อยู่แถว ๆ บ้านนี้) “อยู่หลัง ๆ นาทิดคำ” เป็นต้น

คำบุพบทที่มาจากกริยาคำกริยา ให้ความหมายไม่เจาะจง เช่น “อยู่ละะ ๆ ฮัว” (อยู่แถว ๆ ไร่) เป็นต้น คำวิเศษณ์ที่มาจากกริยาคำกริยาบอกสภาพ ที่ให้ความหมายกริยาเดิม เช่น “ไว ๆ” (เร็ว ๆ) “ซื่อ ๆ” (เฉย ๆ) “แจบ ๆ” (สนิท ๆ) เป็นต้น

คำวิเศษณ์ที่มาจากกริยาคำกริยาและให้ความหมายใหม่ เช่น “ไป ๆ มา ๆ” (ในที่สุด) “อยู่ ๆ” (ทันใด) เป็นต้น

คำวิเศษณ์ที่มาจากกริยาคำนามและให้ความหมายใหม่ เช่น “เว้าหมา ๆ” (พูดซุ่ย ๆ) เป็นต้น

คำวิเศษณ์ที่มาจากกริยาคำบุพบท และให้ความหมาย ไม่ชัดเจน ไม่เจาะจง เช่น “เข้าไปใน ๆ แน่” (เข้าไปข้างใน ๆ หน่อย) “วางอยู่ริม ๆ ฟัน” (วางอยู่ริม ๆ โนน) เป็นต้น

2.3.4 ประโยคภาษาไทยถิ่นอีสาน

ประโยค (Sentence) หมายถึง หน่วยในภาษา ประกอบด้วย ประธาน (Subject) และ ภาคแสดง (Predicate) โดยที่ภาคแสดงใช้บ่งบอกว่าประธานเป็นอย่างไร ทำอะไร เป็นหน่วยทางไวยากรณ์ที่ใหญ่กว่า คำ หรือวลี ประโยคเป็นได้ทั้งภาษาพูดและภาษาเขียน และประโยค 1 ประโยคจะแสดงแนวคิดอย่างสมบูรณ์ 1 แนวคิด และเป็นประพจน์ที่สามารถพิสูจน์ว่าเป็นจริงหรือไม่ได้ โดยประโยคภาษาไทยถิ่นอีสาน แบ่งประเภทตามความซับซ้อนของโครงสร้างประโยค (ชนานันท์ ทรงดี, 2554) ได้ดังนี้

2.3.4.1 ประโยคเดี่ยว หรือ ประโยคเรียบง่าย (Simple sentence) หมายถึง ประโยคที่มีกริยาดำเนินเดี่ยว เช่น “แม่ไปตลาด” (แม่ไปตลาด) “น้องสาวหน้า” (น้องสาวหน้า) เป็นต้น

2.3.4.2 ประโยคซับซ้อน (Complicated sentence) หมายถึง ประโยคที่มีกริยาหลายตัว (Multiverbal sentence) ทั้งกริยาแท้และไม่แท้ แบ่งเป็น 3 ประเภท ดังนี้

1) ประโยคกริยาเรียง หมายถึง ประโยคที่มีกริยาหลายตัวเรียงกันต่อกัน ตั้งแต่ 2 คำขึ้นไปโดยไม่มีคำเชื่อมปรากฏระหว่างคำกริยา และมีความหมายร่วมกันหรือทำนองเดียวกันเพียงเหตุการณ์เดียว เช่น “ลาวพวมเฮ็ดก้อยไข่มดแดงกิน” (เขากำลังทำก้อยไข่มดแดงกิน) “อี่แม่ไปหาเก็บเห็ดมากินแลง” (คุณแม่ไปหาเก็บเห็ดมากินตอนเย็น) เป็นต้น

2) ประโยครวม (Compound sentence) หมายถึง ประโยคที่มีตั้งแต่ 2 ประโยคขึ้นไปรวมเข้าเป็นประโยคเดียวกันด้วยคำเชื่อม ได้แก่ “แต่ว่า” (แต่/แต่ว่า) “แต่ว่า...พั้น” (แต่...กลับ) “ແຫ່ລ້ວກະ/ລ່ະກະ” (และ/และก็) “หรือ/หรือ” (หรือ) เช่น “อี่พ่อบอกให้ข้อยไป แต่ว่า อี่แม่พั้นบอกบ่ให้ไป” (คุณพ่อบอกให้ฉันไป แต่ว่าคุณแม่กลับบอกไม่ให้ไป) เป็นต้น

3) ประโยคซ้อน หมายถึง ประโยคที่ประกอบด้วย मुख्यประโยค (ประโยคที่ต้องมีประโยคอื่นมาขยาย) และอนุประโยค ที่เชื่อมเข้าด้วยกันด้วยตัวบ่งชี้ต่าง ๆ เช่น “ข้อยบอกเจ้าจักท้อแล้ว ว่าบ่ให้เฮ็ดจ้งซี่” (ฉันบอกคุณก็ครั้งแล้วว่าไม่ให้ทำอย่างนี้) “ว่า” เป็นตัวบ่งชี้

2.4 ออนโทโลยี

2.4.1 ความหมายและความสำคัญของออนโทโลยี

ออนโทโลยีเป็นศาสตร์ที่กล่าวถึงธรรมชาติของสิ่งที่มีอยู่ และการจัดหมวดหมู่ของสิ่งหนึ่ง ๆ ในขอบเขตความรู้เรื่องใดเรื่องหนึ่งในแนวคิดเดียวกัน ความรู้จะถูกนำมาจัดเป็นหมวดหมู่อย่างมีกฎเกณฑ์อย่างเป็นระบบ และมีเกณฑ์เงื่อนไขแสดงความเป็นจริงเป็นเท็จที่สามารถตอบคำถามของสิ่งที่มีอยู่ได้ (สมชาย ปรากฏเจริญ, 2548)

ออนโทโลยีมีบทบาทสำคัญในการนิยาม หรือบรรยายเชิงความหมายเพื่อความเข้าใจร่วมกันในขอบเขตความรู้เรื่องใดเรื่องหนึ่ง และลดความคลุมเครือของคำศัพท์ (Ushold & Gruninger, 1996) เป็นฐานความรู้ที่สนับสนุนการแลกเปลี่ยน แบ่งปัน และใช้ความรู้ร่วมกัน สามารถใช้งานซ้ำ (Reuse) นำไปอ้างอิงหรืออนุมาน (Inference) เพื่อหาความรู้ใหม่ ๆ หรือนำมาจัดเก็บ ความรู้ ความเชี่ยวชาญ และประสบการณ์ต่าง ๆ ได้ (Noy & McGuinness, 2001) ซึ่งออนโทโลยีที่มีความครบถ้วนสามารถสร้างความเชื่อถือ หรือความเป็นจริงเป็นเท็จตามที่จะเป็น (Reliability) มนุษย์และคอมพิวเตอร์สามารถเข้าใจความหมายตรงกัน ข้อมูลที่จัดเก็บในออนโทโลยีเป็นข้อมูลที่ได้รับการยอมรับร่วมกัน การจัดเก็บความรู้เป็นแบบ โครงร่างพื้นฐานการอธิบายความรู้ ประกอบด้วย บทบาท (Role) ความสัมพันธ์ (Relation) พฤติกรรม และกฎเกณฑ์ (Axiom) ซึ่งทำให้ออนโทโลยีสามารถแสดงข้อมูลข่าวสารได้อย่างถูกต้อง คำศัพท์ที่อยู่ในออนโทโลยีจะได้รับการจัดหมวดหมู่ (Concept) ที่พิจารณาจากคุณลักษณะร่วม (Generalization-Common) และคุณลักษณะพิเศษ (Specialization) การเรียกใช้งานสามารถป้อนคำถามด้วยภาษาธรรมชาติ (Natural Language)

และภาษาเชิงตรรกะ (Descriptive Logic) การพัฒนาและการบำรุงรักษาทำได้ง่าย (สมชาย ปรากฏ เจริญ, 2548) ถูกนำมาใช้ประโยชน์ในงานหลายด้านอย่างกว้างขวาง อาทิ ปัญญาประดิษฐ์ (Artificial Intelligence) เว็บเชิงความหมาย (Semantic web) วิศวกรรมความรู้ (Knowledge engineering) การประมวลผลภาษาธรรมชาติ (Natural language processing) การค้นคืนสารสนเทศ (Information Retrieval) (Alfarries, Aljably, & Al-Razgan, 2017) เป็นต้น

2.4.2 องค์ประกอบของออนโทโลยี

องค์ประกอบทั่วไปของออนโทโลยี (มาลี กาบมาลา, ลำปาง แม่นมาตย์ และ ครรชิต มาลัยวงศ์, 2549; งามนิจ อาจอินทร์, 2557) สรุปได้ดังนี้

2.4.2.1 แนวคิด (Concept) หรือ คลาส (Class) หมายถึง ขอบเขตความรู้เรื่องใดเรื่องหนึ่งที่สนใจ สามารถอธิบายรายละเอียดได้ เช่น คน สัตว์ พืช อาหาร เครื่องดื่ม เป็นต้น

2.4.2.2 อินสแตนส์ (Instances) หรือ ออบเจกต์ (Objects) หมายถึง สิ่งที่ถูกสร้างขึ้นมาและจัดเก็บอยู่ในคลาสต่าง ๆ หรือคำศัพท์ มีความหมายหรือนิยามความหมายไว้ตามออนโทโลยีเรื่องนั้น เช่น “สารบาตร” เป็นอินสแตนส์ของคลาส “ชื่อโรค” หรือ “แขน” “ขา” “ท้อง” เป็นอินสแตนส์ของคลาส “อวัยวะ” เป็นต้น

2.4.2.3 คุณสมบัติ (Property, Slots, Role, Functions) หมายถึง คุณสมบัติต่าง ๆ ที่นำมาใช้อธิบายแนวคิด หรือ คลาส หรือ อินสแตนส์ เช่น คลาสพืชสมุนไพร มีคุณสมบัติ ได้แก่ ชื่อเรียกทั่วไป ชื่อทางวิทยาศาสตร์ ชื่อเรียกในภาษาถิ่นต่าง ๆ ลักษณะสรรพคุณในการรักษาโรค เป็นต้น

2.4.2.4 ความสัมพันธ์ (Relationship) หมายถึง รูปแบบของความสัมพันธ์ระหว่างแนวคิดและคุณสมบัติในลักษณะต่าง ๆ ดังนี้

1) Sub-class หรือ Is-a หรือ ความสัมพันธ์แบบคลาสลูกกับคลาสพ่อแม่ ที่มีคุณสมบัติการถ่ายทอดคุณสมบัติของแนวคิดไปยังคลาสลูก เช่น “พืชวัตถุ” เป็นคลาสลูกของ “ตำรับยาสมุนไพร” หรือ “มนุษย์” และ “สัตว์” เป็นคลาสลูกของ “ผู้ป่วย” เป็นต้น

2) Part-of หรือ ความสัมพันธ์แบบเป็นส่วนหนึ่ง หรือ เป็นส่วนประกอบ เช่น “ดอก” “ผล” “กิ่ง” “ใบ เป็นส่วนประกอบของ “พืชสมุนไพร” เป็นต้น

3) Syn-of หรือ ความสัมพันธ์ที่มีความเหมือนในเชิงความหมายระหว่างกัน สามารถใช้แทนกันได้ เช่น เกสรตัวผู้มีความหมายเดียวกับสมุนไพร

3) Instance-of หรือ ความสัมพันธ์ในลักษณะการเป็นตัวแทน หรือ สมาชิกของแนวคิด เช่น “ใบยานาง” เป็นตัวแทนของ (Instance-of) “พืชวัตถุ” “ทา” เป็นตัวแทนของ “วิธีใช้ยา” เป็นต้น

2.4.2.5 ข้อกำหนดในการสร้างความสัมพันธ์ หรือ แอ็กเซียม (Axioms) หมายถึง ตรรกะ หรือเงื่อนไขสำหรับการแปลความสัมพันธ์ระหว่างแนวคิดกับคุณสมบัติ หรือแนวคิดกับแนวคิด หรือการตรวจสอบคุณสมบัติ เพื่อการแปลความหมายที่ถูกต้อง มีความสำคัญต่อการนำไปอนุมานความรู้เพื่อสร้างความรู้ใหม่

2.4.2.6 ความสัมพันธ์เชิงความหมายอื่น ๆ ซึ่งกำหนดโดยผู้เชี่ยวชาญ

2.4.3 ประเภทของออนโทโลยี

การแบ่งประเภทของออนโทโลยีมีเกณฑ์ในการแบ่งหลายประเภทขึ้นอยู่กับวัตถุประสงค์และความเหมาะสมในการอธิบายข้อมูล ออนโทโลยีจำแนกตามการประยุกต์ใช้ แบ่งได้เป็น 3 ประเภท (สมชาย ปราการเจริญ, 2548) ดังนี้

2.4.2.1 ออนโทโลยีแบบสาธารณะ (Generic Ontology) หรือชื่อเรียกอื่น เช่น ออนโทโลยีระดับบน (Upper Ontology) Core Ontology, Reference Ontology, Top-level Ontology เป็นต้น คือ ออนโทโลยีที่มีโครงสร้างของความรู้ในเชิงกว้าง หรือในเชิงมหภาพ มีความหลากหลายไม่เกี่ยวข้องกับเรื่องใดเรื่องหนึ่ง และมีความเป็นนามธรรมค่อนข้างสูง (Abstract) สามารถนำไปเชื่อมโยงกับความรู้ศาสตร์สาขาอื่น ๆ และนำมาใช้งานซ้ำ (Reuse) ได้โดยไม่ต้องสร้างใหม่ เช่น Cyc-Upper เป็นออนโทโลยีแบบสาธารณะที่อธิบายสิ่งต่าง ๆ โดยแยกเป็นประเภทแบบไม่ลงรายละเอียดย่อย

2.4.2.2 ออนโทโลยีเฉพาะ (Domain Oriented Ontology) คือ ออนโทโลยีที่มีรายละเอียดครอบคลุมความรู้เฉพาะด้านสำหรับระบบงานใดระบบงานหนึ่ง โดยการรวบรวมคุณลักษณะทั่วไปที่สามารถใช้ร่วมกัน (Generalization) ระหว่างกิจกรรมต่าง ๆ (Task) ภายใต้อบบงานนั้น ๆ ตัวอย่างออนโทโลยีเฉพาะ เช่น Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) (El-Sappagh et al., 2018) และ Unified Medical Language System (UMLS) (National Library of Medicine, n.d.) เป็นออนโทโลยีคำศัพท์ด้านการแพทย์ และ TOVE (University of Toronto, n.d.) เป็นออนโทโลยีคำศัพท์ทางธุรกิจ เป็นต้น

2.4.2.3 ออนโทโลยีสำหรับกิจกรรม (Task Oriented Ontology) คือ ออนโทโลยีที่นำเสนอแนวคิดที่มีขอบเขตความรู้เรื่องใดเรื่องหนึ่งแบบเฉพาะเจาะจง (Particular Domain) เพื่อตอบสนองการทำงาน (Task) เฉพาะงานโดยเฉพาะ เช่น การวินิจฉัย การควบคุม การมอบหมายงาน การวางแผน ตัวอย่าง เช่น E.coli ออนโทโลยีอธิบายจีโนมแบคทีเรีย เป็นต้น

2.4.4 กระบวนการพัฒนาออนโทโลยี

การพัฒนาออนโทโลยีสามารถทำได้หลายวิธี เช่น การนำเอาออนโทโลยีที่มีอยู่แล้วมาใช้ การพัฒนาออนโทโลยีจากฐานข้อมูล การพัฒนาจากศัพท์ควบคุมที่มีอยู่แล้ว เช่น ธิซอร์สในสาขาใดสาขาหนึ่ง และการสร้างออนโทโลยีขึ้นมาใหม่ การเลือกใช้แต่ละวิธีขึ้นอยู่กับวัตถุประสงค์ บริบท หรือความพร้อมในการพัฒนา

การพัฒนาออนโทโลยีขึ้นมาใหม่โดยส่วนใหญ่จะใช้แนวคิดตามวงจรการพัฒนาออนโทโลยี (Life cycle for building ontology) ของ Uschold 4 ขั้นตอน ได้แก่ (1) การกำหนดวัตถุประสงค์ ขอบเขต ประเภทของออนโทโลยีที่จะสร้าง และการนำออนโทโลยีไปใช้ให้ชัดเจน (2) การสร้างออนโทโลยี โดยใช้วิธีนิยามแนวคิด 4 แนวคิด ได้แก่ ระดับกว้างไปสู่แนวคิดเฉพาะเจาะจง (Top-Down) วิธีนิยามแนวคิดที่เฉพาะเจาะจงไปหาแนวคิดระดับกว้าง (Bottom-Up) วิธีนิยามแนวคิดแบบศูนย์กลาง (Middle-Out) และวิธีนิยามแนวคิดเชื่อมโยง (Combination) ที่เป็นแนวคิดผสมผสานระหว่างแบบการนิยามแนวคิดแบบกว้างไปสู่แนวคิดเฉพาะเจาะจงกับการนิยามแนวคิดเฉพาะเจาะจงไปหาแนวคิดในระดับกว้าง (3) การประเมินความถูกต้อง ความน่าเชื่อถือ และการนำกลับมาใช้ใหม่ และ (4) การจัดทำคู่มือการใช้และการพัฒนาออนโทโลยี (มาลี กาบมาลา, ลำปาง แม่นมาตย์ และ ครรชิต มาลัยวงศ์, 2549)

นอกจากนี้ยังมีแนวคิดการพัฒนาแบบอื่น ๆ ที่พัฒนาขึ้นมาเพื่อให้เหมาะสมกับการพัฒนาออนโทโลยีเฉพาะแต่ละโดเมน เช่น การใช้มาตรฐานทางด้านภูมิศาสตร์เป็นแนวคิดในการพัฒนาออนโทโลยีภูมิศาสตร์ (Hakimpour & Gepper, (2001; 2002; 2005); Hakimpour, 2003) การใช้แนวคิดการพัฒนาโครงการมาใช้กำหนดแนวคิดในการพัฒนาออนโทโลยีสำหรับงานโครงการ (Jones; Bench-Capon, & Visser, 1999) วิธีการวิศวกรรมความรู้ (Knowledge-Engineering Methodology) ที่พัฒนามาจากประสบการณ์ในการพัฒนาออนโทโลยี (Noy & McGuinness, 2001) แนวคิดการพัฒนาออนโทโลยีที่มาจากการใช้แนวคิดฮิวริสติก (Heuristic-based) (Sugumaran & Storey, 2002) และแนวคิดแบบวิศวกรรมออนโทโลยี (Ontology Engineering) (งามนิจ อัจฉินทร์, 2557) เป็นต้น

สำหรับงานวิจัยนี้ได้ทำการศึกษาแนวคิดสำหรับใช้ในการพัฒนาออนโทโลยี คำศัพท์ตำรายาโบราณอีสาน 2 แนวคิด คือ วิธีวิศวกรรมความรู้ (Noy & McGuinness, 2001) และวิศวกรรมออนโทโลยี (งามนิจ อัจฉินทร์, 2557) โดยแต่ละแนวคิดมีขั้นตอนในการดำเนินการ ดังนี้

2.4.4.1 การพัฒนาออนโทโลยีด้วยวิธีการวิศวกรรมความรู้ (Noy & McGuinness, 2001) แบ่งเป็น 7 ขั้นตอน ดังนี้

1) กำหนดโดเมนและขอบเขตของออนโทโลยี (Determine the Domain and Scope of the Ontology) เป็นขั้นตอนของการกำหนดโดเมน (Domain) ของออนโทโลยี และตรวจสอบขอบเขตของออนโทโลยีว่าต้องเกี่ยวข้องกับอะไรบ้าง โดยการรวบรวมคำถาม คำตอบต่าง ๆ ที่เกี่ยวข้องกับขอบเขตงานที่กำลังศึกษา การระดมความคิดจากผู้เชี่ยวชาญหรือบุคคลที่เกี่ยวข้องเพื่อหาคำตอบ (Term) ที่เกี่ยวข้องเพื่อนำไปใช้แทนแนวคิด (Concept) โดยคำถามพื้นฐานสำหรับใช้เป็นคำถามเพื่อกำหนดโดเมนและขอบเขตของออนโทโลยีที่จะสร้าง ได้แก่ โดเมนของออนโทโลยีคืออะไร ออนโทโลยีใช้ตอบคำถามอะไรบ้าง คำตอบที่ออนโทโลยีจะต้องเตรียมไว้ตอบคำถามคืออะไร และใครเป็นผู้ใช้และดูแลออนโทโลยี

2) พิจารณาเลือกใช้ออนโทโลยีที่มีอยู่แล้ว (Consider Reusing Existing Ontologies) หรือการนำเอาออนโทโลยีที่มีผู้พัฒนาเอาไว้แล้วมาปรับใช้ให้เหมาะสมกับขอบเขตที่ศึกษา โดยการกำหนดกลุ่มเป้าหมายเพื่อสร้างความเข้าใจเกี่ยวกับความต้องการของผู้ใช้ออนโทโลยี วิเคราะห์กลุ่มเป้าหมายและสร้างแนวคิดของความรู้ที่สนใจ นำไปเปรียบเทียบกับออนโทโลยีที่มีอยู่แล้ว แล้วนำเอาแนวคิดหรือคลาสที่เกี่ยวข้องมาใช้

3) กำหนดคำศัพท์หรือนิยามสำคัญของศัพท์ในออนโทโลยี (Enumerate Important Terms in the Ontology) คือ การพิจารณาแจกแจงรายละเอียดของคำศัพท์ในออนโทโลยีที่สนใจ เพื่ออธิบายให้ผู้พัฒนาและผู้ใช้งานเข้าใจ โดยการเขียนคำศัพท์ที่เป็นไปได้เกี่ยวกับสิ่งที่สนใจศึกษาพร้อมระบุคุณสมบัติของคำศัพท์เหล่านั้น โดยละเอียด

4) กำหนดคลาสและลำดับของคลาส (Define Classes and the Class Hierarchy) เป็นขั้นตอนการระบุคลาสและโครงสร้างของคลาส รวมถึงกำหนดคุณสมบัติของคลาส จำนวนค่า ชนิดของค่า และค่าที่เป็นไปได้ของคุณสมบัติของคลาส โดยวิธีต่าง ๆ ได้แก่ การพัฒนาแบบบนลงล่าง (Top-Down) โดยเป็นการกำหนดแนวคิดทั่วไปของเรื่องที่น่าสนใจ และลำดับของแนวคิดก่อน แล้วจึงแบ่งหมวดหมู่ของคลาส การพัฒนาแบบล่างขึ้นบน (Bottom-Up) เป็นการระบุคลาสส่วนใหญ่ ก่อนแล้วจึงนำมาจัดกลุ่มให้เป็นแนวคิดใหญ่ และการพัฒนาแบบผสม (Combination) เป็นการผสมผสานระหว่างวิธีการ พัฒนาแบบบนลงล่างและวิธีการแบบล่างขึ้นบน โดยจะทำเฉพาะแนวคิดที่สำคัญก่อน แล้วจึงจัดหมวดหมู่ของคลาส

5) กำหนดคุณสมบัติของคลาส (Define the Properties of Classes-slots) คือ การกำหนดคุณลักษณะหรือความสัมพันธ์ระหว่างคลาส เพื่ออธิบายโครงสร้างของคลาสว่ามีความสัมพันธ์ระหว่างกันอย่างไร

6) กำหนดข้อจำกัดของคุณสมบัติ (Define the Facets of the Slots) เป็นขั้นตอนการกำหนดเงื่อนไขหรือข้อจำกัดต่าง ๆ ของสล็อต (Slots) ซึ่งจะเป็นคุณสมบัติของคลาส

หรือ ความสัมพันธ์ (Relation) ระหว่างคลาส ได้แก่ การกำหนดจำนวนค่าของสล็อต (Slot Cardinality) ชนิดของค่าของสล็อต (Slot-value Type) และการกำหนดโดเมนและช่วงของสล็อต (Domain and Range of a Slot)

7) สร้างตัวอย่างของข้อมูล (Create Instances) เป็นการกำหนดข้อมูลในคลาสแต่ละลำดับชั้น

2.4.4.2 การพัฒนาออนโทโลยีด้วยวิธีวิศวกรรมออนโทโลยี (งามนิจ อาจอินทร์, 2557) แบ่งเป็น 6 ขั้นตอน ดังนี้

1) ตรวจสอบโดเมนและขอบเขตของออนโทโลยี ได้แก่ การกำหนดโดเมนและขอบเขตของออนโทโลยีที่จะสร้างว่าเกี่ยวข้องกับสิ่งใดบ้าง โดยจะต้องตอบคำถามได้ว่าโดเมนของออนโทโลยีคืออะไร ออนโทโลยีสร้างขึ้นเพื่อตอบคำถามอะไร และออนโทโลยีเกี่ยวข้องกับสิ่งใดบ้างภายใต้โดเมนที่สนใจ

2) จัดกลุ่มสิ่งที่ยังเกี่ยวข้องในออนโทโลยี ได้แก่ การจำแนกสิ่งต่าง ๆ ที่อยู่ในออนโทโลยีนี้ (Terms) แยกเป็นประเภทคำนาม (noun) คำกริยา (verb)

3) กำหนดคลาส (Class) ตัวอย่างของข้อมูล (Instances) ของแต่ละคลาสและความสัมพันธ์ระหว่างคลาส ได้แก่ การนำคำนามที่มีความหมายเหมือนกันมาจัดกลุ่มเป็นคลาสและกำหนดความสัมพันธ์ระหว่างคลาสในแบบลำดับชั้น (Hierarchy) แบ่งเป็น 2 ลักษณะ ได้แก่ superClassOf และ subClassOf

4) กำหนดคุณสมบัติ (Properties) ให้กับคลาส ได้แก่ การนำคำกริยาที่ได้จัดกลุ่มไว้ในขั้นตอนที่ 2 มากำหนดเป็นคุณสมบัติของคลาสแต่ละคลาส

5) การจำแนกประเภทคุณสมบัติ ได้แก่ การจำแนกประเภทของคุณสมบัติว่าจัดเก็บข้อมูลในลักษณะใด แบ่งเป็น 2 ประเภท ได้แก่ “ObjectProperty” ใช้เก็บความสัมพันธ์ระหว่างคลาสกับคลาส หรือระหว่างตัวอย่างข้อมูลกับตัวอย่างข้อมูล ประกอบด้วย การกำหนดโดเมนและขอบเขตของ ObjectProperty และการสืบทอดคุณสมบัติจากคลาสไปยังตัวอย่างข้อมูล (Instances) และ “DatatypeProperty” เป็นคุณสมบัติที่ใช้จัดเก็บความสัมพันธ์ระหว่างคลาส หรือตัวอย่างข้อมูล กับค่าของข้อมูลของคุณสมบัติแต่ละชนิด ที่อยู่ในรูปแบบ Literal โดยชนิดของข้อมูลที่จัดเก็บได้แก่ ข้อความ (String) ตัวเลข (Integer) หรืออยู่ในรูปแบบ XML Schema datatypes ที่ประกอบด้วย การกำหนดโดเมนและขอบเขตของ DatatypeProperty และการสืบทอดคุณสมบัติจากคลาสไปยังตัวอย่างข้อมูล (Instances)

6) การกำหนดเงื่อนไขให้กับคุณสมบัติแต่ละตัว ได้แก่ การกำหนดเงื่อนไข (Facet) ให้กับคุณสมบัติแต่ละตัว ได้แก่ จำนวนข้อมูลของคุณสมบัติแต่ละตัวที่จะมีได้ (Cardinality) และข้อจำกัดของข้อมูล (Range Restriction)

7) การกำหนดความสัมพันธ์แบบ subPropertyOf ได้แก่ การสร้างความสัมพันธ์แบบลำดับชั้นระหว่างคุณสมบัติของคลาสด้วยตัวเอง

แนวคิดในการพัฒนาออนโทโลยีทั้งสองแนวคิด ส่วนใหญ่มีขั้นตอนการดำเนินการคล้ายคลึงกัน แต่การจัดลำดับการดำเนินการบางขั้นตอนอาจมีความแตกต่างกันบ้าง เช่น การพัฒนาออนโทโลยีด้วยแนวคิดวิศวกรรมความรู้ จะกำหนดให้ขั้นตอนการสร้างตัวอย่างข้อมูลอยู่ในขั้นตอนสุดท้าย ในขณะที่แนวคิดแบบวิศวกรรมออนโทโลยีกำหนดไว้ในขั้นตอนแรก ๆ เป็นต้น

2.4.5 ภาษาที่ใช้ในการพัฒนาออนโทโลยี

ภาษาที่ใช้ในการพัฒนาออนโทโลยี คือ ภาษาสำหรับนำเสนอออนโทโลยีเพื่อให้คอมพิวเตอร์สามารถประมวลผลนิยามศัพท์ที่สร้างขึ้นมาได้ แบ่งประเภทตามรูปแบบภาษาเป็น 3 ประเภท ได้แก่ Logic based, Frame based และ Web based (Guarino, 1998) ดังนี้

2.4.4.1 Logic based คือ ภาษาที่ใช้ตรรกะในการอนุมานความรู้ ประกอบด้วยภาคแสดง (Predicate) อาร์กิวเมนต์ (Argument) และตัวบอกปริมาณ (Quantifier) ตัวอย่างภาษาเช่น ภาษา CYCL, CLASSIC, LOOM เป็นต้น

2.4.4.2 Frame based คือ ภาษาอธิบายออนโทโลยีที่ใช้หลักการปฏิสัมพันธ์ของมนุษย์ เมื่อนึกถึงสิ่งใดจะเชื่อมโยงคุณลักษณะเด่นของสิ่งนั้นด้วย ประกอบด้วยคลาส (Classes) หรือเฟรม (Frame) คุณสมบัติ (Attribute) หรือสล็อต (Slot) และฟาเซต (Facets) ตัวอย่างภาษาเช่น ภาษา OKBC, F-logic เป็นต้น

2.4.5.3 Web based คือ ภาษาอธิบายออนโทโลยี ที่พัฒนามาจากภาษาสำหรับอธิบายทรัพยากรบนเว็บ ใช้หลักการของภาษาประเภทที่ 1 และ 2 อธิบายออนโทโลยีเข้าใจได้ง่าย ตัวอย่างภาษาในกลุ่มนี้ เช่น XML, RDF, OWL เป็นต้น โดยแต่ละภาษามีลักษณะสำคัญ (งามนิจ อัจฉรินทร์, 2557) ดังนี้

1) XML (Extensible Markup Language) เป็นภาษาสำหรับการอธิบายทรัพยากรบนเว็บที่สามารถแลกเปลี่ยนข้อมูลผ่านระบบอินเทอร์เน็ต หรือคอมพิวเตอร์ที่มีระบบแตกต่างกัน อาทิ มีระบบปฏิบัติการ หรืออยู่บนโปรแกรมประยุกต์ต่างกัน โครงสร้างเอกสาร XML ประกอบด้วย 2 ส่วนหลัก คือ ส่วนหัวของเอกสาร (Prolog) ใช้สำหรับอธิบายเวอร์ชันของเอกสาร และอักษรที่ใช้ในเอกสาร และส่วนเนื้อหาทั้งหมด (Root Element) การอธิบายข้อมูลแต่ละอติเมนต์

(Element) ประกอบด้วย แท็กเปิด “<” แท็กปิด “</>” และข้อมูลที่อยู่ภายในแท็กเปิดและปิด ผู้ใช้สามารถสร้างและจัดโครงสร้างของเอกสาร XML ได้เอง

2) RDF (Resource Description Framework) เป็นภาษามาตรฐานที่ใช้อธิบายลักษณะและความสัมพันธ์ของข้อมูลที่คอมพิวเตอร์หรือโปรแกรมประยุกต์สามารถอ่านและเข้าใจได้ด้วยตัวเอง หรือสามารถแลกเปลี่ยนข้อมูลระหว่างคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการและแอปพลิเคชันต่างกันได้ RDF มีโครงสร้างและรูปแบบพื้นฐานเหมือนภาษา XML (Extensible Markup Language) ที่ได้มีการเพิ่มเติมคุณลักษณะให้สามารถอธิบายความหมายและความสัมพันธ์ของข้อมูลได้มากกว่าภาษา XML โดยโครงสร้างของภาษาแบ่งเป็น 3 รูปแบบ ได้แก่ แบบกราฟแบบกลุ่มของ Triple Statements และแบบ XML

3) RDF Schema หรือ RDFS เป็นภาษาที่ใช้นิยามหรือกำหนดโครงสร้างของ RDF โดยอธิบายความหมายของคลาสและคุณสมบัติของคลาส ความสัมพันธ์ระหว่างคลาส ความสัมพันธ์ระหว่างคุณสมบัติ รวมถึงขอบเขต (Domain) และขนาด (Range) ของคุณสมบัติ RDFS เมื่อนำมาใช้ร่วมกัน RDF จะทำให้คอมพิวเตอร์สามารถอนุมาน (Inference) ข้อมูลในระดับที่ลึกกว่าการใช้ RDF แต่การใช้ RDFS ยังมีข้อจำกัด เนื่องจาก RDFS ที่ประกอบด้วยคลาสและคุณสมบัติของคลาส ซึ่งการแสดงคุณสมบัติของคลาสมีเพียง 4 ประเภท ได้แก่ subClassOf, Type, Range และ Domain นั้น ยังไม่เพียงพอต่อการบรรยายข้อมูลให้มีความหมายเชิงตรรกะได้

4) ภาษา OWL (Web Ontology Language) เป็นภาษาอธิบายออนโทโลยีที่พัฒนาขึ้นมาเพื่อแก้ไขข้อจำกัดของเอกสาร RDFS ได้แก่ การเพิ่มคำอธิบายความสัมพันธ์ระหว่างคลาสและความสัมพันธ์ระหว่างคุณสมบัติ (Properties) กำหนดให้คลาสลูก (Subclass) สามารถมีขอบเขตหรือช่วงข้อมูล (Range) แตกต่างจากคลาสแม่ได้ คลาสแต่ละคลาสสามารถสร้างความสัมพันธ์นอกเหนือจากความสัมพันธ์แบบ subClassOf ได้ เช่น นำข้อมูล 2 คลาสมารวมกัน (Union) สมาชิกในคลาสหนึ่งต้องไม่ไปอยู่ในอีกคลาสหนึ่ง (Disjoint) ได้ และความสัมพันธ์แบบเท่าเทียม (Equivalent) เช่น คลาส 2 คลาสเป็นข้อมูลประเภทเดียวกัน เป็นต้น หรือการกำหนดคุณสมบัติพิเศษอื่น ๆ ให้กับคุณสมบัติ เช่น Transitive Property, Invers Property หรือ Functional Property เป็นต้น ทำให้ OWL มีความสามารถในการอนุมานได้ลึกยิ่งขึ้น โครงสร้างการอธิบายข้อมูลที่เป็นรูปแบบง่าย ๆ และซับซ้อนมากขึ้น 3 รูปแบบ ได้แก่ OWL Lite, OWL DL และ OWL Full

2.4.6 โปรแกรมสร้างออนโทโลยี

ตัวอย่างบางส่วนของโปรแกรมที่นำมาใช้สร้างออนโทโลยี มีดังนี้

2.4.6.1 Hozo-Ontology Editor พัฒนาโดยมหาวิทยาลัยโอซากา (Osaka University) ประเทศญี่ปุ่น เป็นโปรแกรมสนับสนุนการพัฒนาออนโทโลยี (Ontology editor) ตามทฤษฎีพื้นฐานประกอบไปด้วย Ontology editor, Onto-Studio และ Ontology server สามารถแสดงผลได้หลายรูปแบบ อาทิ LISP, Text, XML/DTD เป็นต้น และสนับสนุนการนำกลับมาใช้ใหม่ (ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2555; Control Device Division Enegate Co, Ltd., 2011)

2.4.6.2 Protégé พัฒนาโดยมหาวิทยาลัยแอสตันฟอร์ด เป็นโปรแกรมประเภทเปิดเผยแพร่โค้ดต้นฉบับ (Open source) ส่วนติดต่อกับผู้ใช้เป็นแบบกราฟิก (Graphic User Interface : GUI) มีรุ่นสำหรับติดตั้งใช้งานบนเครื่องคอมพิวเตอร์ (Desk top version) และใช้งานบนเว็บเบราว์เซอร์ สามารถใช้สร้างออนโทโลยีได้หลากหลายรูปแบบ เช่น RDF, OWL เป็นต้น และมีโปรแกรมเสริมที่ช่วยสนับสนุนการสร้างออนโทโลยี เช่น OntoGraf, SPARQL, Reasoner (Stanford University, n.d.) เป็นต้น

2.4.6.3 NeOn Toolkit เป็นโปรแกรมประเภทเปิดเผยแพร่โค้ดต้นฉบับ ของ NeOn Technologies Foundation มีมุมมองนำเสนอออนโทโลยีเป็นแบบกราฟ สนับสนุนการพัฒนาออนโทโลยีแบบ OWL, RDF สามารถเชื่อมต่อกับโปรแกรมเสริมอื่น ๆ เช่น SPARQL, Reasoner แต่ไม่สนับสนุนการสร้างกฎ (Rule language) (Erdmann & Waterfeld, 2012)

2.4.6.4 TopBraid Composer พัฒนาโดยบริษัท TopQuadrant Inc. เป็นโปรแกรมสร้างออนโทโลยีประเภทเว็บแอปพลิเคชัน มีเครื่องมือสนับสนุนการพัฒนาแอปพลิเคชันในแบบ Dynamic ontology-driven ตั้งแต่ขั้นตอนการพัฒนาจนถึงการใช้งานตามวงจรการพัฒนาแอปพลิเคชันเชิงความหมาย (Semantic Application Lifecycle) สนับสนุนการแลกเปลี่ยนข้อมูลแบบ RDF, OWL การสืบค้นข้อมูลด้วย SPARQL, Reasoner และการใช้กฎ (TopQuadrant, Inc.2021, www)

การเลือกใช้โปรแกรมสำหรับพัฒนาออนโทโลยี ขึ้นอยู่กับหลายปัจจัย เช่น งบประมาณ ข้อจำกัดของโปรแกรม ภาษาพัฒนาออนโทโลยีที่รองรับ การใช้งานง่าย ไม่ซับซ้อน เป็นต้น ซึ่งในงานวิจัยนี้ เลือกใช้โปรแกรม Protégé รุ่นติดตั้งและใช้งานบนเครื่องคอมพิวเตอร์ส่วนบุคคล โดยเกณฑ์ในการเลือก ได้แก่ สามารถใช้สร้างออนโทโลยีโดเมนใดก็ได้ ไม่เฉพาะเจาะจงโดเมนใดโดเมนหนึ่งเท่านั้น รองรับภาษาออนโทโลยีหลายภาษา สามารถนำเข้าและนำออกข้อมูลได้หลายภาษา ส่วนติดต่อกับผู้ใช้เป็นแบบกราฟิกง่ายต่อการใช้งาน และมีคู่มือ หรือคั่นหาวิธีแก้ปัญหาจากอินเทอร์เน็ตได้ง่าย เป็นต้น

2.4.7 การประเมินผลออนโทโลยี

วิธีการประเมินผลออนโทโลยีทำได้หลายวิธี จำแนกได้ ดังนี้

2.4.7.1 วิธีประเมินผลแบ่งจำแนกตามประเภทและวัตถุประสงค์ของออนโทโลยี แบ่งเป็น 4 กลุ่ม (Brank, Grobelnik, & Mladenic, 2009) ดังนี้

1) การเปรียบเทียบกับมาตรฐานที่ดีที่สุด (Golden standard) ซึ่งอาจจะเป็นออนโทโลยีที่มีอยู่แล้ว หรือค่าสถิติที่คำนวณได้จากคลังข้อความ (Corpus) หรือเตรียมโดยผู้เชี่ยวชาญ โดยสมมติที่ว่าทั้งสองออนโทโลยีมีตัวอย่างข้อมูล (Instances) แบบเดียวกัน และประเมินผลโดยการวัดความคล้ายคลึงกัน

2) การประเมินจากการใช้ออนโทโลยีผ่านโปรแกรมประยุกต์ (Application-based Evaluation) และประเมินผลลัพธ์ที่ได้ เนื่องจากการใช้ออนโทโลยีอาจอยู่รูปของโปรแกรมประยุกต์หรือเป็นงานหนึ่ง (Task) และประสิทธิภาพของโปรแกรมประยุกต์หรืองานเหล่านั้นจึงขึ้นอยู่กับออนโทโลยีที่ใช้ ดังนั้นออนโทโลยีที่ดีจะช่วยให้โปรแกรมประยุกต์มีผลลัพธ์ที่ดีเช่นเดียวกัน

3) การเปรียบเทียบกับข้อมูลที่เป็นโดเมนเดียวกันกับออนโทโลยีที่ถูกประเมิน (Data-driven Evaluation) เช่น การนำแนวคิดและความสัมพันธ์ในออนโทโลยีไปจัดประเภทด้วยโมเดลจำแนกประเภท (Classification) ที่ได้จากการเรียนรู้ของเครื่อง หรือการหาจำนวนแนวคิดที่เหมือนกัน (Overlap) ระหว่างออนโทโลยีและคลังข้อความ (Corpus) เป็นต้น

4) การประเมินโดยมนุษย์ (Assessment by Human) ซึ่งเป็นผู้เชี่ยวชาญในความรู้แขนงนั้น ๆ ว่าออนโทโลยีมีลักษณะเป็นไปตามหลักเกณฑ์มาตรฐาน และความต้องการที่ได้มีการกำหนดเอาไว้เพียงใด

2.4.7.2 วิธีการประเมินผลจำแนกตามระดับของการประเมิน แบ่งเป็น 6 ระดับ (Brank, Grobelnik, & Mladenic, 2009) ดังนี้

1) ระดับคำ (Lexicon) คำศัพท์ (Vocabulary) หรือชั้นข้อมูล (Data Layer) เป็นการประเมินที่ให้ความสำคัญกับแนวคิด (Concept) ตัวอย่างข้อมูล (Instant) ข้อเท็จจริง (Facts) และอื่น ๆ ในออนโทโลยี รวมถึงคำศัพท์ที่เป็นตัวแทน หรือระบุความหมายของแนวคิดที่อยู่ในออนโทโลยี การประเมินใช้วิธีวัดความคล้ายคลึงกันของข้อความ (String) ระหว่างออนโทโลยีที่ต้องการประเมินกับออนโทโลยีที่เป็นมาตรฐาน (Golden standard) เป็นต้น

2) ระดับโครงสร้างแบบลำดับชั้น (Hierarchy) หรืออนุกรมวิธาน (Taxonomy) เป็นการประเมินที่ให้ความสำคัญกับความสัมพันธ์ระหว่างแนวคิดแบบ is-a เป็นอันดับแรก และความสัมพันธ์ในลักษณะอื่น ๆ เช่น คำที่มีความหมายคล้ายกัน เป็นต้น โดยการเปรียบเทียบส่วนที่ซ้อนทับกันของอนุกรมวิธาน (Taxonomic overlap) ของออนโทโลยีที่สร้างขึ้นกับออนโทโลยีมาตรฐาน หรือคลังข้อความ

3) ความสัมพันธ์เชิงความหมายในลักษณะอื่น (Other Semantic Relations) เป็นการประเมินความสัมพันธ์ในลักษณะอื่นของออนโทโลยีที่นอกเหนือความสัมพันธ์แบบ is a เช่น การวัดค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) จากการสืบค้นข้อมูล เป็นต้น

4) ระดับบริบทหรือแอปพลิเคชัน (Context or Application level) เนื่องจากออนโทโลยีอาจเป็นส่วนหนึ่งของระบบสารสนเทศขนาดใหญ่ ซึ่งอาจมีการอ้างอิงไปยังหลายออนโทโลยี หรือถูกอ้างอิงจากออนโทโลยีอื่น การประเมินผลออนโทโลยีในกรณีนี้จึงจำเป็นต้องคำนึงถึงบริบทในการนำไปใช้ในลักษณะดังกล่าวด้วย ดังนั้นการประเมินผลออนโทโลยีจึงพิจารณาจากผลลัพธ์ของแอปพลิเคชันที่ได้มีการนำออนโทโลยีนี้นมาใช้

5) ระดับวากยสัมพันธ์ (Syntactic Level) หรือระดับประโยค โดยปกติออนโทโลยีมักถูกอธิบายด้วยภาษาที่เป็นทางการ และประโยคต้องตรงกับหลักวากยสัมพันธ์หรือไวยากรณ์ของภาษาเหล่านั้น รวมถึงไวยากรณ์อื่น ๆ เช่น การนำเสนอของเอกสารด้วยภาษาธรรมชาติ และการหลีกเลี่ยงการซ้ำซ้อน (Loops) ระหว่างคำนิยาม เป็นต้น

6) ระดับโครงสร้าง (Structure) สถาปัตยกรรม (Architecture) และการออกแบบ (Design) เพื่อให้ออนโทโลยีเป็นไปตามหลักการออกแบบ หรือเกณฑ์ที่กำหนดเอาไว้ก่อนหน้าที่จะมีการพัฒนาออนโทโลยี โดยการประเมินเพื่อตรวจสอบว่าออนโทโลยีเป็นไปตามข้อกำหนดเชิงโครงสร้าง สถาปัตยกรรมและมาตรฐานที่ออกแบบไว้หรือไม่ โดยอาศัยผู้เชี่ยวชาญด้านออนโทโลยีหรือผู้เชี่ยวชาญในสาขานั้น ซึ่งจะช่วยให้ได้ออนโทโลยีที่มีความเหมาะสมสำหรับการนำไปปรับปรุงพัฒนาต่อไปในอนาคต

2.4.7.3 การประเมินออนโทโลยีจำแนกตามคุณภาพและความถูกต้องของออนโทโลยี แบ่งเป็น 7 ประเภท (Raad & Cruz, 2015) ดังนี้

1) ความถูกต้อง (Accuracy) วัดจากความถูกต้องของการให้คำนิยามรายละเอียด และคุณสมบัติของแนวคิด

2) ความสมบูรณ์ (Completeness) วัดจากความครอบคลุมประเด็นปัญหาที่สนใจได้อย่างเหมาะสม

3) ความกระชับ (Conciseness) วัดจากองค์ประกอบที่ไม่เกี่ยวข้องในออนโทโลยี หรือความหมายที่ซ้ำซ้อนของคำศัพท์ที่ใช้

4) การปรับใช้ (Adaptability) วัดจากการคาดการณ์การผลการนำออนโทโลยีไปใช้ในสถานการณ์จริง

5) ความชัดเจน (Clarity) วัดจากความชัดเจนในการสื่อความหมายของคำศัพท์ที่ถูกสร้างขึ้น

6) ประสิทธิภาพทางคอมพิวเตอร์ (Computational efficiency) วัดจากความสามารถในการนำเครื่องมือมาใช้ร่วมกับออนโทโลยีในการนำไปใช้งานที่แตกต่างกัน โดยวัดความเข้ากันได้กับเครื่องมือ และความรวดเร็วในการทำงาน

7) ความสอดคล้อง (Consistency) วัดจากภายในออนโทโลยีต้องไม่มีข้อขัดแย้งเกิดขึ้น หรือมีความไม่สอดคล้องระหว่างองค์ประกอบและรายละเอียดในออนโทโลยี

การพัฒนาออนโทโลยีแต่ละประเภทมีวัตถุประสงค์การนำไปใช้ที่แตกต่างกัน และวิธีการประเมินผลมีอยู่หลายวิธีดังกล่าวข้างต้น การพิจารณาเลือกใช้วิธีประเมินผลจึงขึ้นอยู่กับเหมาะสมและความสอดคล้องกับบริบทของการนำไปใช้ ซึ่งในงานวิจัยนี้ได้พิจารณานำเอาวิธีการประเมินความถูกต้อง (Accuracy) ของผลลัพธ์ที่ได้จากการใช้ออนโทโลยีผ่านการใช้โปรแกรมประยุกต์ (Application-based Evaluation) มาใช้เพื่อประเมินประสิทธิภาพของออนโทโลยีที่จะพัฒนาขึ้นมาในงานวิจัยนี้

2.5 คลังศัพท์

2.5.1 ความหมาย

คลังศัพท์ (Lexicon) หรือ ศัพท์านุกรม เป็นศัพท์บัญญัติราชบัณฑิตยสถาน มีชื่อเรียกอื่น ๆ เช่น พจนานุกรม ปทานุกรม หรือ อภิธานคำศัพท์ หมายถึง หนังสือสำหรับค้นคว้าความหมายของคำที่เรียงตามลำดับบท (พจนานุกรมอิเล็กทรอนิกส์ไทย-อังกฤษเล็กจิตรอน เวอร์ชัน 3.0, ม.ป.ป.) ในการอธิบายลักษณะคำแต่ละคำในคลังศัพท์จะประกอบด้วย ความหมายของคำศัพท์ สัทอักษร (Phonetics) ไวยากรณ์ (Syntax) สัญฐานวิทยา (Morphology) เป็นต้น ทั้งนี้จะขึ้นอยู่กับวัตถุประสงค์ในการสร้างคลังศัพท์ (Hirst, 2009)

คลังศัพท์เป็นฐานความรู้ที่มีความสำคัญต่องานการศึกษาและวิจัยทางด้านภาษาศาสตร์เป็นอย่างมาก โดยเฉพาะการนำมาใช้ในงานด้านการประมวลผลภาษาธรรมชาติที่ต้องใช้อ้างอิงคลังศัพท์ในการประมวลผล (Lexicon-based) เช่น การตัดคำ การวิเคราะห์ความรู้สึก (Sentiment analysis) การแปลภาษา การศึกษาด้านภาษาศาสตร์ หรือการใช้อ้างอิง และอธิบายความหมายของศัพท์ (Intasorn, Gertphol, & Sammapun, 2021; Azizan, Jamal, Abdullah, Mohamad & Khairudin, 2019) เป็นต้น

2.5.2 การรวบรวมคำศัพท์เพื่อสร้างคลังศัพท์

วิธีรวบรวมคำศัพท์เพื่อสร้างคลังศัพท์ แบ่งตามแหล่งข้อมูลที่น่ามาใช้คัดเลือกคำศัพท์ แบ่งเป็น 2 วิธี ได้แก่ การรวบรวมจากพจนานุกรม และการรวบรวมจากคลังคำ (อิสรภาพ ลีรัตนไชยวงศ์, 2560) ดังนี้

2.5.2.1 การรวบรวมจากพจนานุกรม (Dictionary-based) เป็นการสร้างคลังศัพท์จากคลังศัพท์ หรือพจนานุกรม ที่มีอยู่แล้ว เช่น WordNet พจนานุกรมฉบับราชบัณฑิตยสถาน พจนานุกรมอิเล็กทรอนิกส์ไทย-อังกฤษ เล็กซิตรอน (LEXiTRON) (Sornlertlamvanich, 2008) ตัวอย่างคลังศัพท์ที่สร้างโดยรวมคำศัพท์จากพจนานุกรม เช่น Thai WordNet สร้างจากพจนานุกรมภาษาอังกฤษ-ภาษาไทย และเครือข่ายคำ Princeton WordNet (Thoongsup et. al, 2009) คลังศัพท์บอกความรู้สึก สร้างจากพจนานุกรมเล็กซิตรอน (Intasorn, Gertphol & Sammapun, 2021) และพจนานุกรมออนไลน์ Longdo Dict สร้างจากพจนานุกรม ประกอบด้วย Hope Dictionary, Nontri Dictionary, คลังศัพท์ไทย (ThaiGlossary) พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542 และ ศัพท์บัญญัติราชบัณฑิตยสถาน (Metamedia Technology, n.d.) เป็นต้น

2.5.2.2 การรวบรวมจากคลังคำ (Corpus-bases) คลังคำเป็นคลังข้อมูลภาษาเขียนหรือภาษาพูดที่เป็นภาษาที่ใช้อยู่จริง เป็นคลังข้อมูลภาษาขนาดใหญ่ สร้างจากการรวบรวมการใช้ภาษาจากแหล่งต่าง ๆ เช่น หนังสือ หนังสือพิมพ์ วารสาร นิตยสาร บทความ บทวิจารณ์ วรรณกรรมสุนทรพจน์ ข้อเขียนที่ปรากฏอยู่บนบนอินเทอร์เน็ต และสื่อสังคมออนไลน์ มาจัดเก็บเอาไว้ ซึ่งการสร้างคลังศัพท์จากคลังคำทำได้ 2 วิธี ได้แก่ (1) การรวบรวมและสร้างคลังคำขึ้นมาใหม่ เช่นคลังศัพท์บอกความรู้สึกในภาษาไทยจากบทวิจารณ์ออนไลน์ สร้างจากคลังคำที่รวบรวมขึ้นมาใหม่จากบทวิจารณ์ภาษาไทย ได้แก่ บทวิจารณ์โรงแรม ภาพยนต์ และบทวิจารณ์แอปพลิเคชันโทรศัพท์มือถือ จำนวน 4,736 บทวิจารณ์ และได้หาคำบอกความรู้สึกสำหรับนำไปใช้วิเคราะห์ความรู้สึกในช่วงเวลาและลบ 112 ชุด (อิสรภาพ ลีรัตนไชยวงศ์, 2560) และ (2) สร้างจากคลังคำที่มีอยู่แล้ว เช่น พจนานุกรมภาษาอังกฤษ Longman สร้างจากคลังคำ Longman Corpus Network ขนาด 330 ล้านคำ พจนานุกรม Macmillan สร้างจาก World English Corpus ขนาด 200 ล้านคำ คลังศัพท์วิชาการทางภาษาอังกฤษธุรกิจ (อังคณา ทองพูน พัฒนาสร, 2560) สร้างจากคลังคำ KKU-BE Corpus ขนาด 16 ล้านคำ เป็นต้น โดยการสร้างคลังศัพท์จากพจนานุกรมมีข้อดีคือ คำศัพท์พร้อมรายละเอียด และคำอธิบายคำศัพท์ จะถูกรวบรวมเอาไว้เรียบร้อยแล้ว ทำให้ง่ายต่อการเลือกไปใช้ได้ทันที แต่มีข้อจำกัด คือ คำศัพท์เฉพาะด้านอาจจะน้อย ในขณะที่การสร้างคลังศัพท์จากคลังคำจะทำให้ได้คำศัพท์จำนวนมากที่มาจากภาษาที่ใช้อยู่จริง ทันสมัย และตรงกับความต้องการ แต่มีข้อเสียคือ การรวบรวมข้อมูลจะมีความยุ่งยากและใช้เวลาในการสร้างมากกว่า และต้องใช้งบประมาณผลภาษารธรรมชาติเข้ามาช่วย เช่น การตัดคำ การกำกับชนิดของคำ และการกำหนดเกณฑ์การเลือกคำศัพท์มาใส่ในคลัง เป็นต้น

สำหรับการพัฒนาคลังศัพท์ภาษาไทยโบราณและคำศัพท์ภาษาไทยถิ่น พบว่า ส่วนใหญ่ใช้วิธีรวบรวมคำศัพท์จากผู้เชี่ยวชาญ ตัวอย่างเช่น พจนานุกรมภาคอีสาน-ภาคกลาง ฉบับปณิธาน สมเด็จพระมหาธีรวงศ์ (ติสุสมหาเถระ) (สมเด็จพระมหาธีรวงศ์, 2515) สารานุกรมภาคอีสาน-ไทย-อังกฤษ (ปรีชา พิณทอง, 2532) รวบรวมคำศัพท์จากความรู้ความเชี่ยวชาญด้านภาษาศาสตร์ของผู้สร้างพจนานุกรม คำศัพท์ในคลังศัพท์ส่วนมากเป็นหมวดคำทั่วไป จำนวนคำศัพท์ในแต่ละเล่มมีประมาณ 15,000 คำ และอีกวิธีหนึ่งคือ รวบรวมคำศัพท์จากเอกสารโบราณประเภทต่าง ๆ อาทิ จารึก จดหมายเหตุ พงศาวดาร กฎหมายโบราณ วรรณคดี ตัวอย่างเช่น พจนานุกรมโบราณศัพท์ ฉบับราชบัณฑิตยสภา (สำนักราชบัณฑิต, 2563) และประมวลคำศัพท์ภาษาลาวในเอกสารโบราณ (สำนักหอสมุดแห่งชาติ, 2545) โดยลักษณะการอธิบายคำศัพท์จะมุ่งเน้นการอธิบายความหมายและการตีความหมายของคำศัพท์ พร้อมบอกรายละเอียดเอกสารหลักฐานที่บอกที่มาของคำศัพท์ และชนิดของคำตามหลักไวยากรณ์

2.5.3 ขั้นตอนการสร้างคลังศัพท์

การสร้างคลังศัพท์ประกอบด้วยขั้นตอนพื้นฐาน 5 ขั้นตอน (พิชญ์นาฏ สุทธิสมณ์, 2557) ดังนี้

2.5.3.1 การวางแผน (Planning) ได้แก่ การประเมินความต้องการ ระบุกลุ่มผู้ใช้ และระดับของผู้ใช้คลังศัพท์ วัตถุประสงค์ของการจัดทำ เตรียมงบประมาณ ศึกษาเปรียบเทียบข้อดีข้อเสียของคลังศัพท์หรือพจนานุกรมอื่น ๆ ที่มีอยู่ และที่กำลังจะสร้าง กำหนดประเภทและคุณลักษณะของคลังศัพท์ที่จะสร้าง (จำนวนคำ หมวดคำศัพท์ คุณสมบัติของคำ) กำหนดกลุ่มคนทำงาน เช่น ผู้เชี่ยวชาญเฉพาะสาขา ผู้ปฏิบัติงาน เป็นต้น ระยะเวลาในการสร้าง การคัดเลือกคำศัพท์ และจัดทำคู่มือปฏิบัติงาน ระบุแหล่งค้นคว้าอ้างอิงที่เป็นมาตรฐาน และการศึกษาเรื่องลิขสิทธิ์ในการนำเอาคำอธิบาย คำแปล

2.5.3.2 การคัดเลือกและเก็บรวบรวมข้อมูล (Data Collection and Selection) ได้แก่ กำหนดเกณฑ์การคัดเลือกคำศัพท์หลัก จัดเตรียมและรวบรวมคลังข้อมูลหรือแหล่งข้อมูลที่มีคำศัพท์ที่ต้องการ เช่น หนังสือ บทความจากแหล่งต่าง ๆ เช่น นิตยสาร เว็บไซต์ กำหนดปริมาณของคำศัพท์และชนิดของข้อมูลคำศัพท์ที่ต้องการ โดยพิจารณาร่วมกับวัตถุประสงค์และประเภทของคลังศัพท์ แล้วเลือกคำศัพท์บรรจุลงในบัญชีคำศัพท์ ซึ่งอาจใช้วิธีแบบอัตโนมัติเลือกมาจากคลังคำ (Corpus) ด้วยโปรแกรมคอมพิวเตอร์ เช่น Concordance หรือวิธีทำด้วยมือ

2.5.3.3 บันทึกและเรียบเรียงพจนานุกรม (Writing) ได้แก่ การให้คำอธิบายความหมายของคำศัพท์ และการตรวจสอบความถูกต้องของคำศัพท์และคำอธิบาย

2.5.3.4 การสร้างคลังศัพท์และทดสอบ (Producing and Testing) ได้แก่ การจัดพิมพ์ หรือสร้างคลังศัพท์ในรูปแบบที่ต้องการ เช่น คลังศัพท์แบบออนไลน์ และทดสอบใช้งาน ได้แก่ การทดลองใช้ค้นหาคำศัพท์ หรือการนำไปประยุกต์ใช้ในด้านต่าง ๆ เป็นต้น

2.5.3.5 การปรับปรุง (Revising) ได้แก่ การนำเอาข้อผิดพลาดที่เกิดขึ้นหลังการนำไปใช้ระยะหนึ่งแล้วมาปรับปรุงคลังศัพท์ให้มีความถูกต้องหรือผิดพลาดน้อยลง

2.5.4 คลังศัพท์อิเล็กทรอนิกส์

คลังศัพท์อิเล็กทรอนิกส์ หรือพจนานุกรมอิเล็กทรอนิกส์ เป็นฐานความรู้ที่ทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติ และมีบทบาทสำคัญต่อการแปลภาษาด้วยเครื่อง คลังศัพท์อิเล็กทรอนิกส์มีความแตกต่างจากพจนานุกรมทั่วไปคือ รายละเอียดของคำจะอยู่ในรูปสัญลักษณ์ (Symbols) และรหัส (Code) การอธิบายความหมายของคำแต่ละคำจะมีความหมายเดียวไม่กำกวม เพื่อใช้แทนคำอธิบายเกี่ยวกับ คำ ความหมาย โครงสร้าง และข้อมูลอื่นอีกเป็นจำนวนมาก ที่มุ่งเน้นให้เครื่องสามารถเข้าใจ ในขณะที่คลังศัพท์ทั่วไปจะอธิบายโดยใช้ภาษาธรรมชาติที่ผู้ใช้สามารถเข้าใจและแปลความได้เอง (Sornlertlamvanich, 1995)

2.5.4.1 ประเภทของคลังศัพท์อิเล็กทรอนิกส์ แบ่งตามการใช้งานได้เป็น 3 ชนิด ดังนี้

- 1) Basic-term Dictionary คือ คลังศัพท์ที่บรรจุคำศัพท์หมวดทั่วไปทุกสาขา มีคำไม่น้อยกว่า 100,000 คำ
- 2) Technical-term Dictionary หรือ Terminological Dictionary คือ คลังศัพท์ที่บรรจุคำศัพท์เฉพาะด้าน เฉพาะสาขา เช่น แพทย์ คอมพิวเตอร์ กฎหมาย เป็นต้น
- 3) User Specific Dictionary หรือ User Specifics Term Banks คือ คลังศัพท์ที่สร้างขึ้นมาที่มีจุดมุ่งหมายในการใช้งานพิเศษโดยเฉพาะ โดยการดึงข้อมูลบางส่วนออกมาจากคลังศัพท์อื่น หรือเพิ่มข้อมูลพิเศษบางประการเข้าไป

2.5.4.2 องค์ประกอบของคลังศัพท์อิเล็กทรอนิกส์ (Sornlertlamvanich, 1995) ประกอบด้วยองค์ประกอบ ดังนี้

- 1) ข้อมูลสำหรับวิเคราะห์คำ เพื่อให้คอมพิวเตอร์เข้าใจคำ เช่น ชนิดของคำ (Part of speech) ความหมายของคำศัพท์ (Definition) เป็นต้น
- 2) ข้อมูลสำหรับวิเคราะห์ประโยค เพื่อรู้ตำแหน่งและความสัมพันธ์ของคำในระดับโครงสร้าง เช่น ชนิดของคำ

3) ข้อมูลสำหรับวิเคราะห์ความหมายของคำ เพื่อให้เข้าใจความหมายของประโยค เช่น ประเภทของคำ (a-kind-of) คำพ้องความหมาย (Synonym) คำตรงข้าม (Antonym) ส่วนประกอบ (Meronym) และพ้องเสียง (Homonym) เป็นต้น

4) ข้อมูลที่เกี่ยวข้องกับคำในภาษาอื่น ที่มีความหมายตรงกัน เท่ากัน เช่น คำของกลุ่มภาษาที่ต้องการแปล เช่น คู่คำศัพท์ภาษาไทยกับคำศัพท์ภาษาอังกฤษ คำศัพท์ภาษาไทยถิ่นอีสาน-ภาษาไทยมาตรฐาน เป็นต้น

2.5.4.3 ลักษณะการอธิบายคำศัพท์

การอธิบายคำศัพท์ในคลังศัพท์อิเล็กทรอนิกส์ ประกอบด้วยข้อมูลหลายรายการ ขึ้นอยู่กับวัตถุประสงค์ของการพัฒนาคลังศัพท์ เช่น รายการคำศัพท์ (Entry word) ชนิดของคำ (Part of speech) การจัดกลุ่มความหมาย (A-Kind-Of) สัทวิทยา (Morphological) คำที่สะกดคำในแบบอื่น (Different spelling) หรือ ตัวย่อ (Abbreviation) คำที่ใช้เฉพาะท้องถิ่น (Geographical regions used) คำผันหรือคำที่เปลี่ยนรูปตามไวยากรณ์ (Inflections) การออกเสียงคำ (Pronunciations) ความถี่ในการใช้คำหรือความสำคัญของคำในโครงสร้าง (Frequency count, Degree of importance) บอกลักษณะของการใช้คำ เช่น ทางการ/ไม่เป็นทางการ เก่า/ใหม่ ภาษาพูด/ภาษาเขียน ภาษาโบราณ เป็นคำศัพท์เฉพาะสาขา โครงสร้างและตัวอย่างการใช้ สำนวน (Idioms, Idiomatic Expressions) การเกิดร่วมของคำในประโยค (Collocations) การจัดกลุ่มคำตามความหมายเดียวกัน (Thesaurus Information) แหล่งข้อมูลหรือการอ้างอิงถึงเนื้อหาข้อมูล (Information Source) (Sornlertlamvanich, 1995)

ตัวอย่างเช่น ลักษณะการอธิบายคำศัพท์ในพจนานุกรมอิเล็กทรอนิกส์แบบหลายภาษา MMT (Multilingual Machine Translation Dictionary) ที่มีวัตถุประสงค์เพื่อใช้ในการแปลภาษาด้วยเครื่อง (Charoenporn, Kruengkrai, Theeramunkong & Sornlertlamvanich, 2006) ประกอบด้วย (1) ข้อมูลทางภาษาศาสตร์ ได้แก่ สัทวิทยา (Morphological) ไวยากรณ์ (Syntactic) และความหมาย สร้างตามทฤษฎีทางภาษาศาสตร์ (2) รายละเอียดข้อมูลที่จำเป็นสำหรับการแก้ความกำกวมและการเลือกคำแปลที่เหมาะสมในการแปลภาษา ได้แก่ ชนิดของคำแบ่งเป็น 13 ชนิดหลัก และ 51 ชนิดคำย่อย ที่สนับสนุนการวิเคราะห์วากยสัมพันธ์และเป็นตัวแทนกฎไวยากรณ์ (Grammar rule representation) (3) ข้อมูลแปลงไวยากรณ์กับความสัมพันธ์เชิงความหมาย และแปลงไวยากรณ์กับความหมาย สำหรับกำหนดโครงสร้างไวยากรณ์และความหมาย ได้แก่ การจัดกลุ่มคลาสแบบเชิงความหมาย (semantic class) AKO (A-Kind-Of) สำหรับรายการคำ (Word entry) จะเลือกจากการกำหนดลักษณะความยาวของคำ แทนการเลือกคำสั้น ๆ เพื่อลดความยุ่งยากในการเขียนกฎสำหรับการแปลด้วยเครื่อง เป็นต้น

2.6 การประมวลผลภาษาธรรมชาติ

2.6.1 ความหมายของการประมวลผลภาษาธรรมชาติ

ภาษาธรรมชาติ (Natural Language) หมายถึง ภาษาที่มนุษย์ใช้สื่อสารกัน ในแบบ ภาษาพูดและภาษาเขียน และการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) คือ กระบวนการทำให้คอมพิวเตอร์เข้าใจภาษาที่มนุษย์ใช้สื่อสารกันทั้งภาษาพูดและภาษาเขียน รวมถึงทำให้คอมพิวเตอร์สามารถสื่อสารโต้ตอบกับมนุษย์ได้ด้วยของภาษามนุษย์ NLP เป็นสาขาหนึ่งของ เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) (Charoenporn, Kruengkrai, Theeramunkong & Somlertlamvanich, 2006) ถูกนำมาใช้ประโยชน์ในด้านต่าง ๆ เป็นจำนวนมาก อาทิ การรู้จำเสียงและ คำพูด (Speech Recognition) การแปลภาษาด้วยเครื่อง (Machine Translation) การค้นคืนสารสนเทศ (Information Retrieval) การสรุปใจความสำคัญหรือจับประเด็นสำคัญจากข้อความ การจัดหมวดหมู่ เนื้อหา การวิเคราะห์อารมณ์ ความรู้สึก และความคิดเห็นของผู้คน การตรวจสอบสแปม (Spam) อีเมล สร้างผู้ช่วยเสมือน เช่น Siri Alexa หรือ การแปลงเสียงเป็นข้อความ หรือข้อความเป็นเสียง (Bhadwal, Agrawal & Madaan, 2020) เป็นต้น แบ่งเป็น 2 กลุ่ม คือ การเข้าใจภาษาธรรมชาติ (Natural Language Understanding: NLU) และการสร้างภาษาธรรมชาติ (Natural Language Generation: NLG) เทคนิควิธีที่ใช้ ได้แก่ เทคนิควิธีแบบฐานกฎ (Rule-based System) การเรียนรู้ด้วยเครื่อง (Machine Learning) และการเรียนรู้เชิงลึก (Deep Learning)

2.6.2 กระบวนการทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติ

กระบวนการทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติ (Ghosh & Gunning, 2019) มี ดังนี้

2.6.2.1 การวิเคราะห์สัทฐานวิทยา (Morphological Processing) คือ การวิเคราะห์ องค์ประกอบของหน่วยคำว่าประกอบด้วยองค์ประกอบใดบ้าง เช่น หน่วยคำภาษาอังกฤษ ประกอบด้วย คำเติมหน้าหรือคำอุปสรรค (Prefix) รากศัพท์ (Root) และคำต่อท้ายหรือคำปัจจัย (Suffix) เป็นต้น

2.6.2.2 การวิเคราะห์ความหมายของคำ (Lexicon Analysis) คือ การวิเคราะห์ ความหมายของคำแต่ละคำ

2.6.2.3 การวิเคราะห์ไวยากรณ์ (Syntactic Analysis) คือ การวิเคราะห์โครงสร้าง ความสัมพันธ์ของคำและไวยากรณ์ของประโยค (Grammar) และการกำกับชนิดของคำ (Parsing)

2.6.2.4 การวิเคราะห์ทางความหมาย (Semantic Analysis) คือ การวิเคราะห์ ความหมายของคำและความถูกต้องของประโยค

2.6.2.5 การวิเคราะห์ประโยค (Discourse) คือ การศึกษาความสัมพันธ์ หรือความเชื่อมโยงของประโยค 2 ประโยค ที่การตีความของประโยคหนึ่งขึ้นอยู่กับประโยคหนึ่งก่อนหน้า และส่งผลต่อการตีความในประโยคถัดไป

2.6.2.6 การวิเคราะห์การใช้ภาษา (Pragmatic Analysis) คือ การตีความคำและประโยคโดยอ้างอิงสถานการณ์หรือฐานความรู้เดิมซึ่งอาจไม่ได้ระบุอยู่ในเนื้อหานั้น ๆ คล้ายกับวิธีตีความของมนุษย์ที่มีการเชื่อมโยงความรู้ใหม่เข้ากับความรู้เดิม

2.6.3 ขั้นตอนการประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ (Ghosh, S., and Gunning, D., 2019) มีขั้นตอนดังนี้

2.6.3.1 การตัดคำ (Tokenization) ได้แก่ การนำเอาประโยคมาแยกออกเป็นคำ (Words/Tokens) วิธีการตัดคำแบบคำโดด หรือมีคำเดียว เรียกว่า Unigrams การตัดคำแบบ 2 คำติดกัน เรียก Bigrams การตัดคำแบบ 3 คำติดกัน เรียก Trigrams และตัดคำแบบหลายคำติดกัน เรียก n-grams โดย n คือ จำนวนคำติดกัน ที่ต้องการตัดแยกออกจากกัน ซึ่งการกำหนดขนาดของคำที่จะตัดคำขึ้นอยู่กับวัตถุประสงค์และความต้องการในการนำไปใช้งาน โดยปัญหาของการตัดคำ คือ การตัดคำแต่ละแบบมีความผิดพลาดเกิดขึ้นได้ โดยเฉพาะในคำประเภทกลุ่มคำหรือวลี เช่น ประโยคภาษาอังกฤษ “I am living in The United Stated.” เมื่อนำไปตัดคำแบบคำโดด ผลลัพธ์ที่ได้คือ “I”, “am”, “living”, “in”, “The”, “United”, “Stated.” ทำให้ “United” และ “State” ถูกแยกออกจากกันเป็นคนละคำ ซึ่งเป็นการตัดคำที่ผิด (Ghosh, S., and Gunning, D., 2019) หรือตัวอย่างการตัดคำในประโยคคำรับขานภาษาไทยถิ่นอีสาน เช่น “ยาแก้ปวดท้องให้เอาข่าตำใส่น้ำต้มทาท้อง” หมายถึง การปรุงยาแก้ปวดท้องให้ใส่ “น้ำต้ม” (ใส่น้ำที่ต้มสุกแล้ว) แต่เมื่อตัดคำแบบคำโดด คือ “ยา” “แก้” “ปวด” “ท้อง” “ให้” “เอา” “ข่า” “ตำ” “ใส่” “น้ำ” “ต้ม” “ทา” “ท้อง” ทำให้เกิดการตีความเป็นการนำเอาข่าไปตำใส่ “น้ำ” แล้วจึงเอาไป “ต้ม” ในขณะที่บริบทนี้ควรตัดคำเป็น “น้ำต้ม” เป็นต้น

2.6.3.2 การแจกส่วน (Part Of Speech: POS) ได้แก่ การกำกับหมวดคำหรือชนิดของคำแต่ละคำในประโยค เพื่อใช้สำหรับการวิเคราะห์ห้วงคำประกอบและความสัมพันธ์ระหว่างคำในประโยค ตัวอย่างการกำกับหมวดคำให้กับประโยค เช่น [(‘The’, ‘DT’), (‘sky’, ‘NN’), (‘is’, ‘VBZ’), (‘blue’, ‘ADJ’)] โดยการกำกับคำด้วยแท็ก “DT” หมายถึง คำนำหน้านาม (Determiner) “NN” หมายถึง คำนามทั่วไปเอกพจน์ “VBZ” หมายถึง กริยา रूपปัจจุบันที่มีประธานเป็นเอกพจน์บุรุษที่ 3 และ “ADJ” หมายถึง คำวิเศษณ์ (Ghosh & Gunning, 2019)

2.6.3.2 การกรองคำหยุด (Stop Words Removal) หรือคำฟุ่มเฟือย ที่เมื่อกรองออกไปแล้วจะไม่ส่งผลกระทบต่อความหมายของประโยค ตัวอย่างคำหยุดในภาษาอังกฤษ เช่น a an the also of by at etc. too เป็นต้น หรือ คำหยุดในภาษาไทย เช่น แต่ ซึ่ง อีก ใน ได้ โดย เอง ยัง เมื่อ เป็น รวม จาก ต่อ ถ้า กัน เป็นต้น

2.6.3.3 การทำข้อความให้เป็นบรรทัดฐาน (Text Normalization) เนื่องจากการเรียกชื่อสิ่งต่าง ๆ อาจเรียกหลายแบบ เช่น ชื่อเมือง “มุมไบ” (Mumbai) มีชื่อเรียกอีกชื่อหนึ่งคือ “บอมเบย์” (Bombay) หรือชื่อเมืองหลวงของประเทศไทยมีชื่อเรียกหลายชื่อ เช่น “กรุงเทพมหานคร” “กรุงเทพฯ” “กทม.” “เบงคอก” (Bangkok) “บางกอก” เป็นต้น หรือคำกริยาภาษาอังกฤษจะเปลี่ยนรูปตามประธานและกาล (Tense) ยกตัวอย่างเช่น “does”, “doing” เปลี่ยนรูปมาจาก “do” หรือ “began” “begun” เปลี่ยนรูปมาจาก “begin” หรือ “reading” เปลี่ยนเป็น “read” หรือ “watching” เปลี่ยนเป็น “watch” เป็นต้น ในการประมวลผลภาษาธรรมชาติจำเป็นต้องปรับการสะกดคำที่แตกต่างกันนี้ให้อยู่ในบรรทัดฐานแบบเดียวกันด้วยวิธีต่าง ๆ เช่น การตรวจแก้ตัวสะกด (Spelling Correction) การลดรูปคำศัพท์ (Stemming) โดยการผันคำ (Inflection) ให้อยู่ในรูปทั่วไป หรืออยู่ในรากศัพท์ (Root) หรือการใช้คำตามพจนานุกรม (Lemmatization) เป็นต้น

2.6.3.4 การสกัดนิพจน์ระบุนาม (Named Entities Recognition: NER) มีวัตถุประสงค์เพื่อหานิพจน์เพื่อใช้ระบุนามเฉพาะ หรือชื่อเฉพาะ ในประโยค เช่น ชื่อบุคคล ชื่อสถานที่ ชื่อหน่วยงาน ซึ่งมักเป็นคำที่ไม่ปรากฏในพจนานุกรม

2.6.3.5 คำที่มีความกำกวม (Word Sense Disambiguation) หรือ คำที่มีได้หลายความหมาย ตัวอย่าง เช่น “ฝน” ในประโยค (1) หมายถึง น้ำที่ตกลงมาจากเมฆเป็นเม็ด ๆ และในประโยคที่ (2) หมายถึง ทำให้แหลม ทำให้คม และประโยค (3) หมายถึง ทำให้ละเอียด ทำให้เป็นผง เป็นต้น

วันนี้ฝนตกหนักทั้งวัน (1)

เขาเอามีดมาฝนให้คม (2)

เอารากมะเขือฝนใส่น้ำกิน (3)

2.6.2.6 การหาขอบเขตของประโยค (Sentence Boundary Detection) คือ การหาว่าข้อความที่ต่อเนื่องกันนั้นมีขอบเขตของประโยคแต่ละประโยคอย่างไร เช่น ประโยคภาษาอังกฤษ อาจดูขอบเขตของประโยคได้จากเครื่องหมายจุด (Full Stop) หรือเครื่องหมายอื่น ๆ เช่น การขึ้นย่อหน้าใหม่ เครื่องหมายคำถาม (?) เป็นต้น สำหรับในภาษาอื่นที่ไม่มีเครื่องหมายสำหรับบอกขอบเขตประโยคที่ชัดเจนจะต้องใช้วิธีอื่นที่เหมาะสมแทน เช่น การเขียนคำราชาธิสาณ โบราณมีคำที่บอกขอบเขตการขึ้นต้นหรือและจบประโยค เช่น “จักกล่าว” “ยา...” “ดีแล” “เสียก่อน” เป็นต้น

กระบวนการที่กล่าวมานี้ เป็นกระบวนการเตรียมข้อมูล (Pre-processing) ก่อนที่จะนำไปใช้ในกระบวนการอื่นที่เกี่ยวข้องกับงานประมวลผลภาษาธรรมชาติ เช่น การแปลภาษาด้วยเครื่อง การค้นคืนข้อมูล การสรุปใจความสำคัญ และการจัดหมวดหมู่เนื้อหา ทั้งนี้ จะต้องใช้กระบวนการใดบ้างนั้น ขึ้นอยู่กับลักษณะเฉพาะของภาษาแต่ละภาษา

2.7 การแปลภาษาด้วยเครื่อง

2.7.1 ความหมายของการแปลภาษาด้วยเครื่อง

การแปลภาษาด้วยเครื่อง (Machine Translation) หมายถึง การนำเอาเครื่องคอมพิวเตอร์มาใช้แปลภาษามนุษย์จากภาษาหนึ่งไปเป็นภาษาหนึ่ง โดยรูปแบบข้อมูลนำเข้าอาจอยู่ในรูปของตัวอักษรหรือภาษาเขียน หรืออยู่ในรูปของเสียงหรือภาษาพูด (Abdelaal & Alazzawie, 2020)

2.7.2 ประเภทของระบบแปลภาษาด้วยเครื่อง

จำแนกเป็น 3 ประเภท (Abdelaal & Alazzawie, 2020) ดังนี้

2.7.2.1 ระบบแปลภาษาที่อาศัยกฎไวยากรณ์เป็นฐานในการแปลภาษา (Rule-based Machine Translation: RBMT) เป็นการแปลภาษาที่อาศัยกฎไวยากรณ์ทางภาษาศาสตร์ และการวิเคราะห์วากยสัมพันธ์ (Syntax) ของทั้งภาษาต้นฉบับ (Source Language) และภาษาเป้าหมาย (Target Language) ที่สามารถวิเคราะห์เชิงลึกทั้งในระดับไวยากรณ์ (Grammar) และความหมาย แต่เป็นวิธีที่ต้องใช้ความรู้ทางภาษาศาสตร์เป็นอย่างมาก คำศัพท์ในพจนานุกรมมักไม่ครอบคลุมคำศัพท์ที่จะแปลทั้งหมด และการเขียนกฎให้ครอบคลุมทุกภาษาเป็นเรื่องที่ทำได้ยาก โดยขั้นตอนวิธีในการแปลภาษาในกลุ่มนี้แบ่งเป็น 2 ประเภท ได้แก่

1) การแปลทางตรง (Direct Transition) เป็นการแปลจากภาษาหนึ่งไปเป็นภาษาหนึ่ง โดยใช้พจนานุกรม 2 ภาษา (Word-based Translation หรือ Dictionary-based Translation) ประกอบด้วยกระบวนการค้นหาคำแปลของภาษาต้นฉบับแบบคำต่อคำ และการวิเคราะห์กฎไวยากรณ์ภาษาเป้าหมายสำหรับใช้เป็นกฎสำหรับนำคำแปลมาสร้างประโยคในภาษาเป้าหมาย

2) การแปลทางอ้อม (Indirect Translation) ใช้ในการแปลภาษาแบบหลายภาษา (Multi-lingual Translation) แบ่งเป็น 2 ประเภท คือ (1) การแปลแบบถ่ายโอน (Transfer Translation) ประกอบด้วย 3 กระบวนการย่อย ได้แก่ การวิเคราะห์โครงสร้างพื้นฐานวิทยา (Morphology) วิเคราะห์ความหมาย และวากยสัมพันธ์ (Syntax) ของภาษาต้นฉบับ ผลลัพธ์ที่ได้คือ

ตัวแทนของประโยคภาษาต้นฉบับ การถ่ายโอนจากตัวแทนภาษาต้นฉบับไปเป็นตัวแทนภาษาเป้าหมาย และ การสร้างข้อความภาษาเป้าหมาย และ (2) การแปลโดยใช้ตัวแทนภาษากลาง (Interlingua) หมายถึง การสร้างตัวแทนภาษากลางของภาษาต้นฉบับและภาษาเป้าหมายที่ไม่ขึ้นอยู่กับภาษาใดภาษาหนึ่ง ประกอบด้วยกระบวนการวิเคราะห์และสังเคราะห์ คำ ความหมาย และ วากยสัมพันธ์ (Syntax) หรือของข้อความภาษาต้นฉบับและภาษาเป้าหมาย

2.7.2.2 ระบบแปลภาษาที่อาศัยคลังคำ (Data Driven Approach to Machine Translation: DDMT) เป็นระบบการแปลภาษาที่ใช้คลังคำคู่ขนาน (Parallel Corpus) ขั้นตอนวิธีที่ใช้ในการแปลแบ่งเป็น 2 ประเภท ดังนี้

1) การแปลภาษาโดยใช้สถิติ (Statistical Machine Translation System: SMT) เป็นเทคนิควิธีที่ใช้คลังข้อมูลเทียบบท (Parallel Corpus) และสถิติเพื่อคำนวณหาความน่าจะเป็น (Probability) ของคำแปลภาษาเป้าหมาย โดยประโยคคำแปลที่มีค่าความน่าจะเป็นสูงสุดจะถูกเลือกเป็นคำแปลภาษาเป้าหมาย ข้อดีคือ สามารถสร้างคำแปลที่เหมาะสมได้แม้ว่าประโยคที่ต้องการแปลจะไม่เหมือนกับประโยคใด ๆ ในคลังข้อมูล อย่างไรก็ตาม การแปลโดยใช้สถิติยังไม่สามารถแปลสำนวนและวลีที่ต้องพึ่งพาริบทที่อยู่ห่างไกลออกไปได้ (Lopez, 2008) ขั้นตอนวิธีที่ใช้ในการแปลแบ่งเป็น 3 ขั้นตอนวิธี ได้แก่ การแปลแบบคำต่อคำ (Word-based SMT) การแปลแบบวลี (Phrase-based SMT) การแปลแบบลำดับชั้น (Hierarchical Phrase-based)

2) การแปลแบบอิงตัวอย่าง (Example-based Machine Translation: EBMT) เป็นการแปลโดยแบ่งข้อความออกเป็นส่วน ๆ แล้วนำแต่ละส่วนไปแปลโดยการจับคู่ (Matching) ได้แก่ การแบ่งแยกข้อความเป็นส่วน ๆ และนำไปค้นหาหรือจับคู่กับตัวอย่างคำแปลที่คล้ายคลึงกันในฐานข้อมูลโดยใช้ขั้นตอนวิธีจับคู่ ถ้าส่วนย่อยของข้อความที่ถูกแบ่งแยกสามารถจับคู่กันได้จะถูกนำไปรวมเข้าด้วยกันใหม่ (Adaption) และรวมส่วนย่อยต่าง ๆ ที่เกี่ยวข้องกัน (Recombination) เพื่อสร้างเป็นข้อความภาษาเป้าหมาย

2.7.2.3 ระบบการแปลภาษาแบบผสมผสานเทคนิค (Hybrid Machine Translation) เป็นระบบการแปลภาษาที่นำหลาย ๆ แนวทางมาผสมผสานใช้ร่วมกัน เช่น การแปลภาษาโดยใช้โครงข่ายประสาทเทียม (Neural Machine Translation: NMT) เป็นเทคนิควิธีการแปลภาษาโดยการนำเอาคลังข้อมูลภาษาขนาดใหญ่มาเรียนรู้ด้วยเครื่อง (Machine Learning) โดยจุดเด่นของระบบแปลภาษาแบบ NMT คือ การแปลจากตัวแทนประโยคทั้งหมดในภาษาเป้าหมายแทนการแยกเป็นวลีหรือคำ ตัวอย่างระบบแปลภาษาที่เป็นที่รู้จักกันเป็นอย่างดี คือ Google Translate เป็นต้น

2.8 งานวิจัยที่เกี่ยวข้อง

การพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน เป็นงานด้านการประมวลผลภาษาธรรมชาติ ประเภทอักษรและภาษาไทยโบราณ การศึกษางานวิจัยที่เกี่ยวข้อง แบ่งเป็น 3 ส่วน ได้แก่ การแปลภาษาไทยโบราณด้วยเครื่อง การพัฒนาลัทธิศัพท์อิเล็กทรอนิกส์สำหรับการประมวลผลภาษาธรรมชาติ และการใช้ออนโทโลยีสำหรับการแปลภาษาด้วยเครื่อง ดังนี้

2.8.1 การแปลภาษาไทยโบราณด้วยเครื่อง

การวิจัยด้านการแปลภาษาไทยโบราณด้วยเครื่องและงานที่เกี่ยวข้อง ที่มีการศึกษาและนำเสนอวิธีการแปลเอาไว้ในช่วง 15 ปีที่ผ่านมา ส่วนใหญ่เป็นการศึกษาในอักษรและภาษาไทยโบราณที่บันทึกในเอกสารโบราณทางภาคเหนือและภาคอีสาน ภาคเหนือ ได้แก่ อักษรธรรมล้านนา ภาษาไทยถิ่นเหนือ (ภาษาล้านนา) และภาคอีสาน ได้แก่ อักษรธรรมอีสาน ภาษาไทยถิ่นอีสาน ดังนี้

นิกร ยาพรหม และ พุชยดี ศิริแสงตระกูล (2550) นำเสนอวิธีใช้ข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Transition Networks: ANTs) สำหรับการแปลภาษาล้านนาเป็นภาษาไทยมาตรฐาน วิธีวิจัยประกอบด้วย การสร้างกฎโครงสร้างประโยคสามัญ 3 กฎ กฎโครงสร้างวลี 4 กฎ การสร้างพจนานุกรมภาษาล้านนา-ไทย ขนาด 7,497 คำ การอธิบายคำศัพท์ในพจนานุกรมประกอบด้วย คำล้านนา เครื่องหมายเท่ากับ และส่วนวิเคราะห์ความหมายของคำ ประกอบด้วย ประเภทคำ คำที่ถอดจากภาษาล้านนาเป็นภาษาไทย ส่วนกำกับหมวดคำ ลักษณะประจำคำ ความหมายภาษาอังกฤษ และศัพท์อักษร การจัดเก็บข้อมูลในโครงสร้างแบบทรีย์ ตัดคำโดยใช้วิธีการเทียบคำที่ยาวที่สุด และสร้างกฎปริวรรตอักษรล้านนา 226 กฎ ผลการประเมินประสิทธิภาพในการแปลถูกต้อง ร้อยละ 78.84 โดยความผิดพลาดเกิดขึ้นจากโครงสร้างประโยคสามัญที่ประกอบด้วย ภาคประธาน ภาคแสดง และภาคขยาย ไม่ครอบคลุมรูปแบบประโยคภาษาล้านนาทั้งหมด

นงคณัฐ ไพบูลย์ (2552) นำเสนอการแปลงอักษรธรรมอีสานเป็นคำอ่านภาษาไทยถิ่นอีสานและศัพท์อักษร (Phonetic) โดยใช้พจนานุกรมร่วมกับกฎ (Rule) กระบวนการแปลงอักษรธรรมอีสานเป็นคำอ่านภาษาไทยถิ่นอีสาน ประกอบด้วย การตัดคำแบบคำที่ยาวที่สุด (Longest matching) ในกรณีที่ไม่พบคำอ่านในพจนานุกรมจะใช้กฎที่ได้จากการวิเคราะห์โครงสร้างคำของอักษรธรรมอีสานร่วมกับกฎทางภาษาศาสตร์ ที่ประยุกต์มาจากกฎโครงสร้างคำของ ดวงแก้ว สวามิภักดิ์ และกฎการประสมอักษรเป็นคำของอักษรธรรมอีสาน จำนวน 105 กฎ โครงสร้างพจนานุกรม ประกอบด้วย คำศัพท์อักษรธรรมอีสาน คำอ่านอักษรไทยปัจจุบัน ความหมาย และ

ศัพท์อักษร จัดเก็บคำศัพท์ในโครงสร้างทรี (Tries) แบบไฟล์ข้อความ จำนวนคำศัพท์ในพจนานุกรมมีทั้งสิ้น 10,012 คำ การทดสอบประสิทธิภาพด้วยชุดข้อมูลข้อความอักษรธรรมอีสาน เนื้อหาประเภท นิทานชาดก 2 เรื่อง และพุทธทำนาย 16 ข้อ

โสภณ พรหมโสคา (2553) นำเสนอวิธีแปลงอักษรธรรมอีสานเป็นภาษาไทยโดยอาศัยวากยสัมพันธ์ของภาษาไทยและไวยากรณ์ไม่พึ่งบริบท (Context Free Grammar) การตัดคำและแบ่งแยกพยางค์ใช้ในงานวิจัยนี้ใช้วิธีตัดคำแบบยาวที่สุด กรณีที่ไม่พบคำศัพท์ในพจนานุกรมจะใช้กฎตามโครงสร้างของอักษรธรรมอีสานร่วมกับกฎทางภาษาศาสตร์มาใช้ในการแปลงอักษรธรรมอีสานเป็นคำอักษรไทย ใช้พจนานุกรมคำศัพท์อักษรธรรมอีสาน 6,910 คำ การอธิบายคำศัพท์ในพจนานุกรมประกอบด้วยคำศัพท์อักษรธรรมอีสาน คำอ่านอักษรไทย ความหมาย และศัพท์อักษรจัดเก็บคำศัพท์ในโครงสร้างข้อมูลแบบทรี บันทึกในแบบไฟล์ข้อความ ทดสอบความถูกต้องโดยนำแนวคิดไปใช้แปลงเอกสารที่เขียนด้วยอักษรธรรมอีสาน เนื้อหาหมวดวรรณกรรมพื้นบ้าน และพุทธทำนาย ผลการทดสอบมีความถูกต้องอยู่ในระดับร้อยละ 82.13

ณัฐมีมา สุรเดช (2554) นำเสนอวิธีการวิเคราะห์โครงสร้างคำภาษาล้านนาด้วยเทคนิคการทำเหมืองกราฟ โดยการออกแบบรูปแบบกราฟที่เป็นตัวแทนของคำอักษรธรรมล้านนา 77 รูปแบบ แบ่งเป็น 5 ประเภท ได้แก่ (1) โครงสร้างคำที่ประสมพยัญชนะกับสระ (2) โครงสร้างคำที่มีตัวสะกดเขียนไว้ด้านล่างพยัญชนะต้น (3) โครงสร้างคำที่เขียนตัวสะกดไว้หลังพยัญชนะต้น (4) โครงสร้างคำที่มีพยัญชนะซ้อน และ (5) โครงสร้างคำที่มีตัวไหล และทำการทดสอบประสิทธิภาพของแนวคิดนี้โดยนำชุดข้อมูลประเภทวรรณกรรมอักษรธรรมล้านนา เรื่อง “คร่าวขอประวัติ ครัวอภิไชย ขาวปี” ขนาด 4,008 คำ ไปแปลงให้อยู่ในรูปแบบกราฟตามรูปแบบกราฟที่ได้ออกแบบไว้แล้วนำคำที่อยู่ในรูปแบบของกราฟไปทำการวิเคราะห์โครงสร้างคำที่ปรากฏบ่งชี้ด้วยเทคนิคการทำเหมืองกราฟ ผลลัพธ์ที่ได้แสดงให้เห็นถึงลักษณะโครงสร้างที่ได้มีความสอดคล้องกับรูปแบบโครงสร้างกราฟที่ได้ออกแบบไว้ ผลการวิจัยชี้ให้เห็นว่า คำในภาษาล้านนาสามารถนำมาจัดเก็บในรูปแบบโครงสร้างกราฟและใช้เป็นตัวแทนคำภาษาล้านนาแต่ละประเภทได้ รวมถึงนำไปใช้วิเคราะห์โครงสร้างคำภาษาล้านนาได้

เพิ่มพร ลักขณาวรรณกุล และ พุชยดี ศิริแสงตระกูล (Lakkhanawannakun & Seresangtakul, 2014a) นำเสนอวิธีการแปลงข้อความอักษรธรรมอีสานไปเป็นอักษรและภาษาไทยมาตรฐาน โดยใช้พจนานุกรมอักษรธรรมอีสาน-ภาษาไทยมาตรฐานร่วมกับเทคนิคข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Translation Networks: ATNs) พจนานุกรมที่พัฒนาขึ้นมีคำศัพท์หลักพิมพ์ด้วยอักษรธรรมเป็นภาษาไทยถิ่นอีสานและคำแปลภาษาไทยมาตรฐาน ขนาด 8,000 คำ การอธิบายคำศัพท์ ประกอบด้วยคำศัพท์หลักภาษาไทยถิ่นอีสานอักษรธรรมอีสาน และคำแปลภาษาไทยมาตรฐาน (Thai word) หน่วยเสียง (Phoneme) ชนิดของคำย่อย (Sub part of speech) ความหมาย

ในภาษาไทยมาตรฐาน (Thai meaning) เครื่องหมายพิเศษ ความหมายภาษาอังกฤษแปลจากคำอธิบายภาษาไทยมาตรฐาน และคำอธิบายภาษาไทย การพัฒนาพจนานุกรมมีวัตถุประสงค์เพื่อใช้สำหรับการแปลข้อความภาษาไทยถิ่นอีสานอักษรธรรมอีสานไปเป็นข้อความภาษาไทยมาตรฐาน การแปลใช้วิธีตัดคำแบบยาวที่สุด (Longest matching algorithm) ระบบแปลภาษาประกอบด้วย 4 โมดูลหลัก ได้แก่ กระบวนการก่อนการประมวลผลข้อความอักษรธรรมอีสาน (Isarn Dharma text pre-processing) การวิเคราะห์ประโยค (Sentence analysis) การถอดความ (Transcription) และ การแปล ประเมินประสิทธิภาพการแปลด้วยชุดข้อมูลพุทธทำนาย (คำทำนายของพระพุทธเจ้า) และ ประโยคข้อความทั่วไป ความถูกต้องของการแปลอยู่ในระดับร้อยละ 61.81

เพิ่มพร ลักขณาวรรณกุล และ พุชยดี ศิริแสงตระกูล (Lakkhanawannakun & Seresangtakul, 2014b) ทำการวิจัยเพื่อปรับปรุงประสิทธิภาพการตัดคำอักษรธรรมอีสานและการแปลข้อความที่เขียนด้วยอักษรธรรมอีสานไปเป็นภาษาไทยมาตรฐานให้ประสิทธิภาพสูงขึ้น โดยใช้เทคนิควิธีแบบ ATNs วิเคราะห์และแปลประโยค ช่วยเพิ่มความแม่นยำในการตัดคำโดยการจับคู่พยางค์ที่ยาวที่สุดร่วมกับนิพจน์ระบุนาม เพื่อใช้กำหนดขอบเขตของเอนทิตี การทดสอบประสิทธิภาพกับชุดข้อมูลข้อความอักษรธรรมอีสาน พบว่า ความแม่นยำในการแปลโดยใช้เทคนิคแบบ ATNs ที่ใช้การตัดคำแบบยาวที่สุดร่วมกับนิพจน์ระบุนามอยู่ในระดับร้อยละ 70.61 มีประสิทธิภาพดีกว่าการใช้ ATNs เพียงวิธีเดียวที่อยู่ในระดับร้อยละ 67.99

ทัศนวรรณ ศูนย์กลาง สุนีย์ พงษ์พิณิจิญาญ และ วิภาวดี ม่วงอ้น (2558) นำเสนอสถาปัตยกรรมระบบการแปลภาษา 2 ทาง ระหว่างภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐาน โดยใช้ระบบแปลแบบถ่ายทอด (Transfer Machine Translation) กระบวนการแปลแบ่งเป็น 3 ขั้นตอน ได้แก่ เปลี่ยนรูปแบบภาษาต้นฉบับให้อยู่ในรูปแบบตัวแทนภาษาต้นฉบับ นำตัวแทนภาษาภาษาต้นฉบับไปเปลี่ยนเป็นภาษาเป้าหมาย และวิเคราะห์สร้างภาษาเป้าหมายตามลักษณะที่เหมาะสม โดยใช้กฎการถ่ายทอดและพจนานุกรมสำหรับวิเคราะห์ประโยคต้นฉบับ ในส่วนของคำ (Morphology) วายกสัมพันธ (Syntactical) และความหมาย (Semantic) นำมาสร้างเป็นภาษาเป้าหมายตามโครงสร้างภาษาแบ่งเป็น 2 ส่วน ได้แก่ ส่วนสร้างกฎไวยากรณ์จากโครงสร้างประโยคภาษาไทย และส่วนของการแปลภาษา ใช้กฎไวยากรณ์ 34 กฎ พจนานุกรม 2 ภาษา การสร้างระบบแปลภาษาประกอบด้วย การสร้างคลังคำอ่านภาษาไทยมาตรฐาน-ภาษาไทยถิ่นอีสาน 100 ประโยค (1,895 คำ) พร้อมกำกับหน้าที่ของคำ สำหรับคำแปลที่ไม่มีในพจนานุกรมจะใช้วิธีคำนวณหาค่าระยะแก้ไขที่น้อยที่สุด (Minimal Edit Distance) การวิเคราะห์คำ การค้นหาคำแปล การวิเคราะห์ประโยคการแปลใช้กฎการแปลที่ได้จากการวิเคราะห์โครงสร้างไวยากรณ์ การประเมินประสิทธิภาพการแปลประโยคสนทนาภาษาไทย-อีสาน ที่ใช้ทั่วไปในชีวิตประจำวัน ได้แก่ ประโยคบอกเล่า 25 ประโยค ปฏิเสธ 20 ประโยค คำถาม 15 ประโยค ขอร้อง 10 ประโยค คำสั่ง 10 ประโยค ความถูกต้องในการแปล จากภาษาไทยเป็นภาษาไทยถิ่นอีสานอยู่ในระดับร้อยละ 62.5 และแปลจากภาษาไทยถิ่นอีสานเป็นภาษาไทยถูกต้องร้อยละ 70

2.8.2 การพัฒนาคลังศัพท์อิเล็กทรอนิกส์สำหรับงานประมวลผลภาษาธรรมชาติ

ในการวิจัยและพัฒนาคลังศัพท์อิเล็กทรอนิกส์ทั้งแบบภาษาเดียวและหลายภาษานั้น พบว่ามี 3 ประเภท ได้แก่ คลังศัพท์สำหรับมนุษย์ใช้เพื่อค้นหาคำแปล หรือคำอธิบายความหมายของคำศัพท์ต่าง ๆ เช่น พจนานุกรมฉบับราชบัณฑิตยสถาน พจนานุกรม Longdo เป็นต้น และคลังศัพท์สำหรับงานประมวลผลภาษาธรรมชาติ เช่น การตัดคำ การกำกับชนิดของคำ การแปลภาษา การวิเคราะห์โครงสร้างประโยค การวิเคราะห์อารมณ์ความรู้สึก เป็นต้น สำหรับการพัฒนาคลังศัพท์สำหรับระบบการแปลภาษาไทยโบราณด้วยเครื่อง มีการศึกษาและนำเสนอไว้ ดังนี้

พุชชดี ศิริแสงตระกูล และ นิกร ยาพรม (2552) นำเสนอเทคนิควิธีการใช้พจนานุกรมและข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Transition Networks: ATNs)) ร่วมกับการประยุกต์ใช้ออโตมาตาจำกัดเชิงไม่กำหนด (Nondeterminist Finite Automata: NFAs) เพื่อแก้ปัญหาคำภาษาล้านนาที่ไม่มีในพจนานุกรม วิจัยประกอบด้วยการสร้างพจนานุกรมล้านนา-ไทยขนาด 7,497 คำ สำหรับใช้ในการตัดคำ และวิเคราะห์ประโยคล้านนา ผลการวิจัยได้กฎโครงสร้างประโยค 4 กฎ โครงสร้างวลี 6 กฎ ใช้สำหรับการแปลประโยคและเลือกความหมายของคำ โดยพิจารณาจากบริบทในประโยคสำหรับกรณีที่มีคำศัพท์ที่มีหลายความหมาย ผลการทดสอบกับประโยคอักษรล้านนาแบบตัวพิมพ์ จำนวน 558 ประโยค ผลการประเมินประสิทธิภาพความถูกต้องของการแปลระดับประโยคอยู่ในระดับร้อยละ 83.69 ความผิดพลาดการแปลมีสาเหตุจาก การตัดคำผิด ไม่พบคำศัพท์ในพจนานุกรม โครงสร้างวลีและประโยคไม่สอดคล้องกับโครงสร้างประโยคที่แปล และส่วนมูลฐานและส่วนเสริมพิเศษที่เป็นส่วนประกอบของประโยคไม่สอดคล้องกับรูปประโยคภาษาล้านนาที่ใช้ในการแปล ข้อมูลทดสอบส่วนใหญ่เป็นวรรณกรรม คติโบราณ คำพังเพย และคาถาอาคม มีโครงสร้างประโยคไม่ตรงกับโครงสร้างวลีและประโยคในข่ายงานการเปลี่ยนเพิ่มขยาย ATNs

นนุช ไพบูลย์ และ พุชชดี ศิริแสงตระกูล (Phaiboon & Seresangtakul, 2017) นำเสนอการพัฒนาคลังศัพท์อักษรธรรมอีสานสำหรับการประมวลผลภาษาธรรมชาติ (Isarn Dharma Alphabets Lexicon for Natural Language Processing) ผลการวิจัยได้คลังศัพท์ขนาด 8,178 คำ คำศัพท์รวบรวมจากเอกสารโบราณ และตรวจสอบความถูกต้องโดยผู้เชี่ยวชาญทางภาษา การอธิบายคำศัพท์ในคลังศัพท์ ประกอบด้วย คำศัพท์อักษรธรรมอีสาน คำอ่านของคำอักษรธรรมอีสาน (Phonemes) คำแปลภาษาไทยมาตรฐาน คำศัพท์ภาษาอังกฤษ ชนิดของคำ (Parts of Speech) ชนิดของคำย่อย (Sub-parts of Speech) สัญลักษณ์พิเศษ คำอธิบายภาษาไทยและภาษาอังกฤษ คำศัพท์ถูกจัดเก็บไว้ในโครงสร้างทรี (Trie Structure) นอกจากนี้ ในงานวิจัยนี้ได้สร้างกฎการถ่ายถอดคำอักษรธรรมอีสานเป็นคำอ่านภาษาไทยถิ่นอีสาน เพื่อให้แก้ปัญหาคำที่ไม่รู้จัก (คำศัพท์ที่ไม่มีในพจนานุกรม) ใช้เป็นคำอ่าน และการแปลในขั้นตอนต่อไป และใช้สัญลักษณ์พิเศษเพื่อบอกให้ทราบ

ว่าเป็นที่แปลความหมายได้มากกว่าหนึ่งความหมาย ผลการประเมินความถูกต้องของการถ่ายถอดข้อความอักษรธรรมอีสานเป็นข้อความภาษาไทยมาตรฐาน ใช้ข้อมูลชุดทดสอบประกอบด้วย เนื้อหาหมวด วรรณกรรม ตำรายา และบทกวี/คำพังเพย ค่าเฉลี่ยความถูกต้องร้อยละ 78.75

2.8.3 การใช้ออนโทโลยีในการประมวลผลภาษาธรรมชาติ

ออนโทโลยีถูกนำมาใช้สนับสนุนการประมวลผลภาษาธรรมชาติและ การแปลภาษาด้วยเครื่องในกระบวนการเชิงความหมายในหลายภาษา เช่น คำศัพท์ต้นฉบับที่มีความกำกวมและแปลเป็นภาษาเป้าหมายได้หลายคำ โดยตัวอย่างการนำเสนอผลการศึกษา มีดังนี้

โอคุมูระ และ โอวี (Okumura & Hovy, 1994) นำเสนอวิธีเชื่อมโยงคำศัพท์ภาษาญี่ปุ่นกับแนวคิดในออนโทโลยีด้วยพจนานุกรม 2 ภาษาแบบกึ่งอัตโนมัติ เพื่อใช้สำหรับระบบแปลภาษาญี่ปุ่นเป็นภาษาอังกฤษด้วยเครื่องแบบอัตโนมัติ โดยนำเสนอขั้นตอนวิธี 3 ขั้นตอน คือ การจับคู่คำที่มีความหมายเหมือนกันจากแนวคิดที่เกิดร่วมกันระหว่างแนวคิดในออนโทโลยีกับคำในภาษาอังกฤษที่เกี่ยวข้องมากที่สุด การจับคู่อาร์กิวเมนต์ (Argument Match) ในภาษาญี่ปุ่นกับออนโทโลยี เพื่อเสริมความถูกต้องของการจับคู่ที่เหมือนกันที่เกิดจากไม่มีคำภาษาอังกฤษที่มีความหมายเหมือนกันตั้งแต่ 2 คำขึ้นไป และการจับคู่ตัวอย่าง (Example Match) เพื่อเพิ่มความสมบูรณ์ของ 2 วิธีข้างต้น โดยนำตัวอย่างประโยคในออนโทโลยีมาเปรียบเทียบกับคำอธิบายประโยค (Definition Sentences) ในคลังศัพท์ภาษาญี่ปุ่น-ภาษาอังกฤษ ด้วยการวัดความคล้ายคลึงกันของตัวอย่างทั้งสอง ฐานความรู้ที่ใช้ประกอบด้วยออนโทโลยี พจนานุกรมภาษาญี่ปุ่น พจนานุกรมภาษาญี่ปุ่น-อังกฤษ ออนโทโลยีถูกสร้างด้วยวิธีกึ่งอัตโนมัติจากเครือข่ายคำ WordNet และพจนานุกรมลองแมน (Longman Dictionary of Contemporary English) ใช้สนับสนุนกระบวนการวิเคราะห์ กระบวนการสร้างระบบแปลภาษา และเป็นกรอบการทำงานร่วมกันระหว่างคำศัพท์ในภาษาต่าง ๆ พจนานุกรมภาษาญี่ปุ่น-ภาษาอังกฤษ ใช้สำหรับจับคู่คำศัพท์ภาษาอังกฤษกับภาษาญี่ปุ่นที่มีความหมายสอดคล้องกันตามแนวคิดคู่ภาษา (Bilingual Concept) และคำศัพท์ที่ได้จะเป็นคำศัพท์ที่ปรากฏอยู่ในทั้งคลังศัพท์ภาษาญี่ปุ่น และพจนานุกรมภาษาญี่ปุ่น-ภาษาอังกฤษ การอธิบายลักษณะโครงสร้างคำศัพท์ประกอบด้วย การสะกดคำ/อักษรวิธี สันฐานวิทยา และชนิดของคำ วากยสัมพันธ์ และกฎเกณฑ์เชิงความหมาย (Semantic Constraints) และคำแปลภาษาอังกฤษ (English equivalent) คำพ้องความหมาย (Synonyms) และตัวอย่างประโยค 2 ภาษา การจับคู่ภาษาที่มีแนวคิดเท่ากัน (Equivalent Bilingual Concepts) ด้วยออนโทโลยี แบ่งเป็น 4 กรณี คือ 1:1 หมายถึง ภาษาอังกฤษ 1 คำ เชื่อมโยงกับแนวคิดในออนโทโลยี 1 แนวคิด 1:M หมายถึง คำภาษาอังกฤษ 1 คำ เชื่อมโยงกับแนวคิดในออนโทโลยีได้มากกว่าหนึ่งแนวคิด M:1 หมายถึง คำภาษาอังกฤษหลายคำ เชื่อมโยงกับแนวคิดในออนโทโลยี 1 แนวคิด และแบบ M:N หมายถึง คำภาษาอังกฤษแต่ละคำ

เชื่อมโยงกับแนวคิดในออนโทโลยีได้หลายแนวคิดและแนวคิดในออนโทโลยี 1 แนวคิดเชื่อมโยงกับคำภาษาอังกฤษได้หลายคำ โดยกรณี 1:1 และ M:1 ใช้วิธีเชื่อมโยงแนวคิดแบบง่าย ๆ กรณี M:N วิธีจับคู่คำที่เทียบเท่ากัน และกรณีที่ 1:M ใช้การจับคู่อาร์กิวเมนต์และการจับคู่ตัวอย่าง ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีที่นำเสนอ โดยใช้ข้อมูล คำนาม 980 คำ คำกริยา 860 คำ และคำคุณศัพท์ 520 คำ พบว่า มีประสิทธิภาพมากกว่าร้อยละ 80

โคลเล และคณะ (Kolle, Bhagat, Zade, Dand, & Lifna, 2018) นำเสนอแบบจำลองการใช้ ออนโทโลยีในการระบุคำหลัก (Keyword) ในเอกสารตามโดเมน แทนวิธีแบบเดิมที่แยกคำหลักตามชื่อเรื่อง (Title) แบบจำลองประกอบด้วย 2 องค์ประกอบ ได้แก่ การสร้างออนโทโลยี (Ontology Creation Module) ตามแบบจำลอง HCHIRSIM (Frikh, Djaanfar & Ouhbi, 2011) และการสกัดคำหลัก (Keyword Extraction Module) ตามคุณลักษณะ (Feature-based Approach) กระบวนการประกอบด้วย การนำเอกสารที่ต้องการสกัดคำหลักตามโดเมนเข้าสู่ระบบ จับคู่คำในเอกสารกับตัวแทนแนวคิด (Instances) ในออนโทโลยีที่สร้างขึ้น ให้คะแนนคำหลักตามเกณฑ์ที่กำหนด ได้แก่ ความถี่ของคำในภาษาอังกฤษ และความถี่ที่เกิดในข้อความ น้ำหนักตำแหน่ง ชนิดของคำ จำนวนครั้งที่คำถูกใช้เป็น ประธาน กรรม หรือ Predicate ระยะห่างระหว่างตำแหน่งปัจจุบันและก่อนหน้าที่คำดังกล่าวปรากฏขึ้นมา นำคะแนนที่ได้ไปจัดอันดับ และคำหลักที่อยู่อันดับสูงสุดจะถูกใช้เป็นคำหลักโดเมนเฉพาะ ผลประเมินพบว่า ประสิทธิภาพของแบบจำลองที่นำเสนอเมื่อเปรียบเทียบกับโมเดล HCHIRSIM พบว่า มีความแม่นยำอยู่ระหว่างร้อยละ 80 - 90 ค่าการเรียกคืน (Recall) และ f-measure เพิ่มขึ้นร้อยละ 5

จอร์จ มิลเลอ (George, 1995) นำเสนอการสร้างเครือข่ายคำภาษาอังกฤษ (WordNet) ที่มีลักษณะร่วมกันระหว่างคลังศัพท์แบบดั้งเดิมและทฤษฎี (Thesaurus) เพื่อสนับสนุนงานด้านการประมวลผลภาษาธรรมชาติ ด้วยแนวคิดการรวบรวมคำศัพท์ที่มีความหมายเดียวกันเอาไว้ในกลุ่มเดียวกัน หรือมีโน้ตส์เดียวกัน (Concept) แบ่งชนิดของคำศัพท์ (Syntactic Category) ตามวากยสัมพันธ์ เป็น 5 ชนิด ได้แก่ คำนาม คำกริยา คำคุณศัพท์ และคำวิเศษณ์ จำนวนคำศัพท์ในเครือข่ายคำมีประมาณ 118,000 คำ และมีความหมายแตกต่างกันมากกว่า 90,000 ความหมาย (มากกว่า 166,000 คู่) ร้อยละ 17 เป็นคำที่มีความหมายมากกว่าหนึ่งความหมาย (Polysemous) ร้อยละ 40 เป็นคำที่ปรากฏอยู่ในกลุ่มคำไวพจน์มากกว่าหนึ่งกลุ่มขึ้นไป แบ่งประเภทความสัมพันธ์เชิงความหมาย (Semantic Relations) ระหว่างชนิดของคำ เป็น 6 ประเภท ได้แก่ ความสัมพันธ์แบบไวพจน์ (Synonymy) หรือ คำที่มีความหมายเหมือนกัน เช่น pipe-tube, rise-ascend ความสัมพันธ์แบบตรงข้าม (Antonymy/Opposites) ระหว่างคำคุณศัพท์ และคำวิเศษณ์ เช่น wet-dry, powerful-powerless, rapidly-slowly ความสัมพันธ์แบบจำกลุ่ม-ลูกกลุ่ม (Hyponymy-Subordinate) หรือ ความสัมพันธ์แบบมีคุณสมบัติร่วมกัน หรือเป็นสมาชิกในกลุ่มเดียวกัน ระหว่างคำนาม เช่น sugar maple, maple, tree, plant ความสัมพันธ์แบบเป็นส่วนหนึ่ง (Meronymy) หรือเป็นส่วนประกอบ

ระหว่างคำนาม เช่น brim-hat, gin-martini, ship-fleet ความสัมพันธ์แบบ Toponymy (Manner) ใช้สำหรับแสดงความสัมพันธ์ของหมวดคำกริยาในลักษณะเป็นลูกกลุ่ม เช่น march เป็นกริยาลูกกลุ่มของ walk, whisper เป็นกริยาลูกกลุ่มของ speak ความสัมพันธ์แบบคลุมความ (Entailment) ระหว่างคำกริยา ได้แก่ กริยาที่ทำคล้ายกัน drive-ride หรือกริยาที่ทำตรงข้ามกันเช่น divorce-marry อย่างไรก็ตาม ความสัมพันธ์เชิงความหมายที่มีอยู่ยังไม่เพียงพอสำหรับการใช้สร้างประโยค เครื่องหมายคำนี้ได้รับพัฒนาอย่างต่อเนื่องจนปัจจุบันคือ รุ่น 3.0 ซึ่งมีส่วนประกอบหลัก ดังนี้ (1) กลุ่มคำไวพจน์ คือ กลุ่มคำที่มีความหมายเหมือนกัน ได้แก่ แบบสมบูรณ์หรือเหมือนกันทุกประการ แบบไม่สมบูรณ์ที่คล้ายกันมาก หรือมีความหมายตรงตัวเหมือนกัน แต่มีความหมายแฝงหรือขอบเขตการใช้ไม่เหมือนกัน (2) คำอธิบาย (Gloss) หรือคำจำกัดความหมายของกลุ่มคำไวพจน์ (3) ชนิดของคำทางวากยสัมพันธ์ (Syntactic Category) แบ่งเป็น 4 ประเภท คือ นาม กริยา คุณศัพท์ และวิเศษณ์ จำนวนคำศัพท์ทั้งหมดมีทั้งสิ้น 155,287 คำ แบ่งเป็น คำไวพจน์กลุ่มคำนาม 82,115 กลุ่ม คำกริยา 13,787 กลุ่ม คำคุณศัพท์ 18,156 กลุ่ม และคำวิเศษณ์ 3,621 กลุ่ม

สรุปผลการทบทวนวรรณกรรมด้านการแปลภาษาและการปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรและภาษาไทยโบราณแบบอัตโนมัติด้วยเครื่อง พบว่า กระบวนการแปลภาษาไทยโบราณประกอบด้วย 2 กระบวนการหลัก คือ กระบวนการเตรียมข้อมูลก่อนการประมวลผล (Pre-processing) ได้แก่ การตัดคำ ใช้วิธีตัดคำแบบยาวที่สุด และกระบวนการแปลจากภาษาต้นฉบับ เช่น ภาษาล้านนา ภาษาไทยถิ่นอีสาน เป็นภาษาไทยมาตรฐาน (Lakkhanawannakun & Seresangtakul, 2014a) โดยฐานความรู้ที่ใช้ในการแปลภาษาคือคลังศัพท์ เช่น พจนานุกรมภาษาไทยโบราณ-ภาษาไทยมาตรฐาน ที่มีคำศัพท์หลักเป็นคำอักษรไทยโบราณ เช่น อักษรธรรมอีสาน (Phaiboon & Seresangtakul, 2017) อักษรธรรมล้านนา (พุทธดี ศิริแสงตระกูล และ นิกร ยาทรม, 2552) เป็นต้น จำนวนคำศัพท์ในพจนานุกรมประมาณ 6,000 - 10,000 คำ ระดับความถูกต้องของคำแปลอยู่ระหว่างร้อยละ 60 – 82 รายละเอียดดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 งานวิจัยด้านการปริวรรตเอกสารโบราณของไทยด้วยเครื่อง

ผู้วิจัย	อักษร-ภาษาต้นฉบับ	อักษร-ภาษาเป้าหมาย	เทคนิค	ความถูกต้อง
นิกร ยาทรม และ พุทธดี ศิริแสงตระกูล (2550)	ธรรมล้านนา-ภาษาล้านนา	อักษรไทย-ภาษาไทยมาตรฐาน	<ul style="list-style-type: none"> - พจนานุกรมภาษาล้านนา-ภาษาไทย (7,497 คำ) - ATN (Augmented Transition Networks) - กฎโครงสร้างประโยค - ตัดคำแบบยาวที่สุด 	78.84

ตารางที่ 2.1 งานวิจัยด้านการปริวรรตเอกสารโบราณของไทยด้วยเครื่อง (ต่อ)

ผู้วิจัย	อักษร-ภาษา ต้นฉบับ	อักษร-ภาษา เป้าหมาย	เทคนิค	ความ ถูกต้อง
พูนศักดิ์ ศิริแสงตระกูล และนิกร ยาพรม (2552)	ธรรมล้านนา- ภาษาล้านนา		<ul style="list-style-type: none"> - ATN + Nondeterminist Finite Automata - พจนานุกรมภาษาล้านนา-ไทย 7,497 คำ - ตัดคำแบบคำที่ยาวที่สุด 	83.69
นงคฺนุช ไพบุลย์ (2552)	อักษรธรรม อีสาน- ภาษาไทยถิ่น อีสาน	อักษรไทย-คำ อ่านภาษาไทย ถิ่นอีสาน	<ul style="list-style-type: none"> - พจนานุกรมอักษรธรรม อีสาน 10,012 คำ - กฎโครงสร้างคำของอักษร ธรรมอีสาน 105 กฎ - ตัดคำแบบคำที่ยาวที่สุด 	80.51
โสภณ พรหมโสดา (2553)	อักษรธรรม อีสาน	อักษรไทย- ภาษาไทย มาตรฐาน	<ul style="list-style-type: none"> - พจนานุกรมคำศัพท์อักษร ธรรมอีสาน 6,910 คำ 	82.13
ทัศนวรรณ ศูนย์กลาง สุนีย์ พงษ์พินิจภิญโญ และ วิณาวดี ม่วงอัน (2558)	อักษรไทย- ภาษาไทยถิ่น อีสานและ ภาษาไทย มาตรฐาน	อักษรไทย-คำ อ่านภาษาไทย มาตรฐาน- ภาษาไทยถิ่น อีสาน	<ul style="list-style-type: none"> - การแปลภาษา 2 ทาง ด้วย เทคนิคการแปลแบบ ถ่ายทอด (Transfer Machine Translation) - คลังคำอ่านภาษาไทย มาตรฐาน-ภาษาไทยถิ่น อีสาน 100 ประโยค - คำนวณหาค่าระยะแก้ไขที่ น้อยที่สุด สำหรับคำแปลที่ ไม่มีในพจนานุกรมใช้วิธี - กฎการแปล ใช้ กฎจากการ วิเคราะห์โครงสร้าง ไวยากรณ์ - ประเมินประสิทธิภาพการ แปลด้วยประโยคสนทนา ภาษาไทย-อีสาน ที่ใช้ทั่วไป ในชีวิตประจำวันทั่วไป 80 ประโยค 	62.50 - 70.00
นงนุช ไพบุลย์ และ พูนศักดิ์ ศิริแสงตระกูล (2560)	อักษรธรรม อีสาน- ภาษาไทยถิ่น อีสาน	อักษรไทย- ภาษาไทย มาตรฐาน- ภาษาอังกฤษ	<ul style="list-style-type: none"> - พจนานุกรมคำศัพท์อักษร ธรรมอีสาน-ภาษาไทย มาตรฐาน-คำศัพท์ ภาษาอังกฤษ 8,000 คำ 	-

ตารางที่ 2.1 งานวิจัยด้านการปริวรรตเอกสารโบราณของไทยด้วยเครื่อง (ต่อ)

ผู้วิจัย	อักษร-ภาษา ต้นฉบับ	อักษร-ภาษา เป้าหมาย	เทคนิค	ความ ถูกต้อง
จินตนา พลศรี (2564)	คำปริวรรตของ อักษรไทยน้อย (อักษรไทย)- ภาษาไทยถิ่น อีสาน)	คำปริวรรต ของ อักษรไทย น้อย-คำอ่าน ภาษาไทยถิ่น อีสาน- ภาษาไทย มาตรฐาน	<ul style="list-style-type: none"> - คลังศัพท์ คำปริวรรต-คำอ่าน ภาษาไทยถิ่นอีสาน- คำอ่าน ภาษาไทยมาตรฐาน - ออนโทโลยีคำศัพท์ ตำรายา โบราณอีสาน - คัดคำแบบคำที่ยาวที่สุด 	

เนื่องจากอักษรและภาษาไทยโบราณมีลักษณะเฉพาะที่สำคัญแตกต่างจากอักษรและภาษาไทยปัจจุบันคือ จำนวนตัวอักษรโบราณมีน้อยกว่าอักษรไทยปัจจุบัน สะกดคำโดยไม่มีวรรณยุกต์ คำเดียวกันสะกดคำได้หลายแบบ บันทึกเป็นภาษาถิ่น คำศัพท์บางคำสามารถอ่านและแปลได้หลายแบบหลายความหมาย ขึ้นอยู่กับบริบทของคำที่อยู่รอบ ๆ และเนื้อเรื่องที่อ่าน ทำให้การปริวรรตและแปลเอกสารโบราณด้วยคลังศัพท์เพียงอย่างเดียว ไม่สามารถแปลหรือทำคำอ่านของคำศัพท์อักษรไทยโบราณได้อย่างถูกต้อง ซึ่งในอักษรและภาษายุคปัจจุบัน ได้มีการศึกษาและนำเสนอวิธีการแก้ปัญหาในกรณีนี้ เช่น วิธีการใช้ออนโทโลยีเลือกคำแปลที่มีความเหมาะสม (Kolle, Bhagat, Zade, Dand, & Lifna, 2018) หรือสร้างโมเดลการแปลภาษาด้วยการฝึกฝนให้คอมพิวเตอร์เรียนรู้จากคลังคำขนาดใหญ่ เป็นต้น อย่างไรก็ตาม การสร้างโมเดลการแปลภาษาด้วยการเรียนรู้ด้วยเครื่อง สำหรับกรณีอักษรและภาษาไทยโบราณ พบว่า มีข้อจำกัดที่สำคัญคือ คลังคำที่จะให้นำมาศึกษามีอยู่น้อยมาก (Low Resources) และจำนวนไม่เพียงพอที่จะนำมาฝึกและสร้างโมเดลที่มีประสิทธิภาพได้ ดังนั้นในงานวิจัยนี้จึงนำเสนอกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ที่ใช้ฐานความรู้ประเภทคลังศัพท์และออนโทโลยีคำศัพท์ เพื่อแก้ไขปัญหาคำปริวรรตของอักษรไทยน้อยมีคำอ่านภาษาไทยถิ่นอีสานได้หลายแบบหลายความหมาย ขึ้นอยู่กับบริบท โดยคลังศัพท์ที่จะพัฒนาในงานวิจัยนี้จะเป็นคลังศัพท์แบบคำปริวรรต-คำอ่านภาษาไทยมาตรฐาน-คำอ่านภาษาไทยมาตรฐาน ซึ่งมีคำศัพท์หลักเป็นคำปริวรรตที่ถ่ายทอดจากอักษรไทยน้อย หมวดคำศัพท์เป็นหมวดตำรายาโบราณอีสาน

2.9 กรอบแนวคิดในการวิจัย

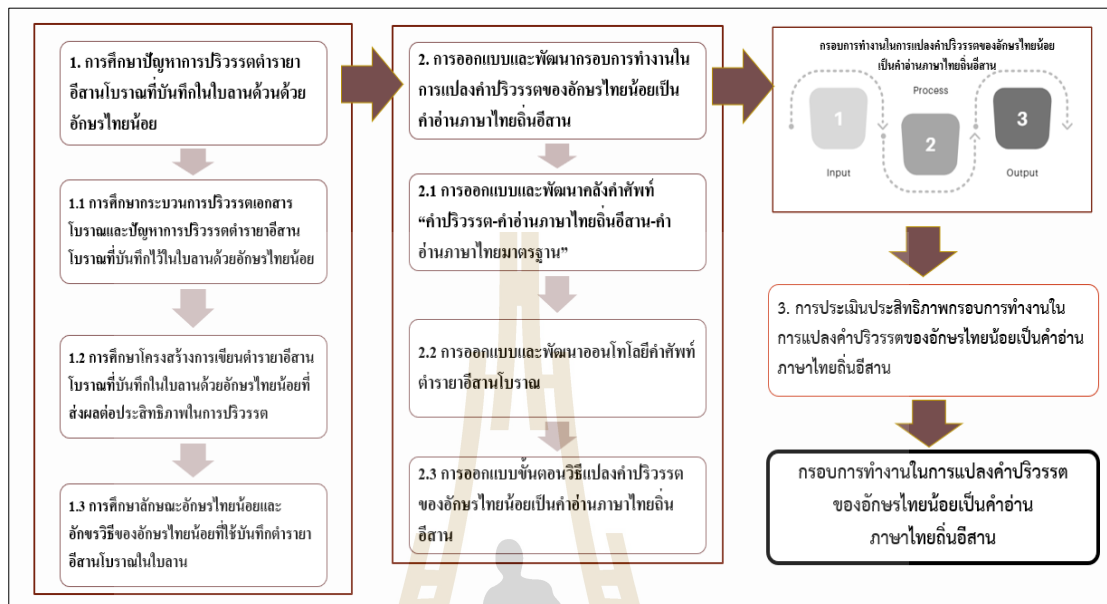
จากการทบทวนวรรณกรรม แนวคิด และทฤษฎีที่เกี่ยวข้อง ได้แก่ การพัฒนาระบบแปลภาษาด้วยเครื่อง การพัฒนาคลังศัพท์อิเล็กทรอนิกส์ การใช้ออนโทโลยีในการประมวลผลภาษาธรรมชาติ รวมถึงการศึกษาลักษณะเฉพาะของอักษรและภาษาไทยโบราณที่เลือกนำมาศึกษาในเบื้องต้น (อักษรไทยน้อยที่ใช้จัดบันทึกคำราชาโบราณอีสานเป็นภาษาไทยถิ่นอีสาน) และสำรวจปริมาณข้อมูลคลังคำคำราชาโบราณอีสานที่จะนำมาศึกษา ในการวิจัยเรื่อง การพัฒนารอบในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน จึงได้กำหนดกรอบแนวคิดในการวิจัยตามขั้นตอนวงจรการพัฒนาระบบ (SDLC) แบ่งเป็น 3 ระยะ ดังนี้

ระยะที่ 1 การศึกษาปัญหาการปริวรรตคำราชาอีสาน โบราณที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย ใช้วิธีวิจัยเชิงคุณภาพ (Qualitative Research) เก็บรวบรวมข้อมูลโดยการสัมภาษณ์แบบกึ่งโครงสร้าง (Semi-structured or Guided Interview) กลุ่มตัวอย่าง คือ นักภาษาภาษาโบราณ และทำการศึกษาโครงสร้างการบันทึกคำราชาโบราณอีสานและอักษรวิธีอักษรไทยน้อยจากเอกสารใบลานที่ได้ปริวรรตและทำคำอ่านเอาไว้แล้ว กลุ่มตัวอย่างคือ คำราชาอีสานที่ปริวรรตจากคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย

ระยะที่ 2 การออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานตามหลักการและทฤษฎี ดังนี้ (1) การพัฒนารอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานตามขั้นตอนการปริวรรตอักษรไทยน้อย (พิมพ์พรรณ ไพบูลย์หวังเจริญ, 2543) (2) การพัฒนาระบบแปลภาษาด้วยเครื่องแบบอาศัยกฎไวยากรณ์เป็นฐานในการแปลภาษา (Rule-based Machine Translation: RBMT) แบบแปลทางตรง (Direct Transition) (Abdelaal & Alazzawie, 2020) (3) การออกแบบและพัฒนาฐานข้อมูลคลังศัพท์ตามขั้นตอนของ ศิริลักษณ์ โรจกิจอำนวย (2552) และ (4) การพัฒนาคลังศัพท์แบบคลังคำ (Corpus) เป็นฐาน (วิโรจน์ อรุณมานะกุล, 2553) และ (5) การพัฒนาออนโทโลยีตามแนวคิดของ นอย และแม็กกินเนส (Noy & McGuinness, 2001)

ระยะที่ 3 การประเมินประสิทธิภาพการพัฒนารอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ใช้วิธีการประเมินผลเชิงปริมาณและเชิงคุณภาพ ได้แก่ การหาค่าร้อยละของระดับความถูกต้อง (Accuracy ของการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน และระดับความเหมาะสมของการพัฒนารอบการทำงานที่นำเสนอ โดยวิธีวัดจากระดับความพึงพอใจ 5 ระดับ (Rating Scale) ตามวิธีของลิเคอร์ท (Likert, 1961)

สรุปรายละเอียดกรอบแนวคิดในการวิจัยดังรูปที่ 2.7



รูปที่ 2.7 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยเรื่อง การพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานนี้ ผู้วิจัยออกแบบการวิจัยโดยแบ่งขั้นตอนการวิจัยเป็น 3 ระยะ ได้แก่ ระยะที่ 1 การศึกษาปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ระยะที่ 2 การออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานและระยะที่ 3 การประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน โดยแต่ละระยะ มีวิธีดำเนินการวิจัยตามลำดับ ดังต่อไปนี้

3.1 ระยะที่ 1 การศึกษาปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย

การศึกษาในระยะที่ 1 มีวัตถุประสงค์เพื่อศึกษาปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย และนำผลการวิเคราะห์มาใช้ในการออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ประกอบด้วยการศึกษา 3 ส่วน ได้แก่ (1) การศึกษากระบวนการปริวรรตเอกสารโบราณและปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย (2) การศึกษาโครงสร้างการบันทึกคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อยที่ส่งผลต่อประสิทธิภาพในการปริวรรต และ (3) การศึกษาลักษณะอักษรไทยน้อยและอักษรวิธอักษรไทยน้อยที่ใช้บันทึกคำราชาโบราณอีสานในใบลาน โดยมีรายละเอียดการวิจัยในแต่ละส่วน ดังนี้

3.1.1 การศึกษากระบวนการปริวรรตเอกสารโบราณและปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย

3.1.1.1 วิธีการวิจัย ใช้วิธีวิจัยเชิงคุณภาพ (Qualitative Research) เก็บรวบรวมข้อมูลโดยการสัมภาษณ์แบบกึ่งโครงสร้าง (Semi-structured or Guided Interview)

3.1.1.2 พื้นที่ในการวิจัย ได้แก่ หน่วยงานซึ่งมีหน้าที่ในการอนุรักษ์และเผยแพร่เนื้อหาในเอกสารโบราณ ที่มีนักภาษาโบราณสังกัดประจำอยู่ในองค์กร และมีผลงานการปริวรรตเอกสารโบราณเผยแพร่อย่างต่อเนื่อง จำนวน 2 แห่ง ได้แก่ กลุ่มหนังสือตัวเขียนและจารึก หอสมุด

แห่งชาติ ทำวาศุกรี กรุงเทพมหานคร และสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม

3.1.1.3 กลุ่มตัวอย่างที่ใช้ในการศึกษา คือ นักภาษาโบราณ ที่สังกัดหน่วยงานที่เลือกเป็นพื้นที่ทำการวิจัย ได้แก่ กลุ่มหนังสือตัวเขียนและจารึก หอสมุดแห่งชาติ ทำวาศุกรี กรุงเทพมหานคร จำนวน 7 คน และสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม จำนวน 5 คน รวมทั้งสิ้น 12 คน

3.1.1.4 การรวบรวมข้อมูลและเครื่องมือในการรวบรวมข้อมูล คือ แบบสัมภาษณ์แบบกึ่งโครงสร้าง กำหนดประเด็นคำถาม 3 ด้าน ดังนี้

1) ข้อมูลส่วนตัว ได้แก่ ประสบการณ์ และความเชี่ยวชาญด้านการปริวรรตเอกสารโบราณ (อักษรและภาษาโบราณที่มีความเชี่ยวชาญ หมวดเนื้อหาที่มีความเชี่ยวชาญ)

2) ปัญหาการปริวรรตเอกสารโบราณ และปัญหาการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย และวิธีแก้ไข

3) ปัญหาอื่น ๆ ที่เกี่ยวข้อง

3.1.1.5 การวิเคราะห์ข้อมูล ได้แก่ สรุปละเอียดในการปริวรรตเอกสารโบราณ และการปริวรรตคำราชาโบราณอีสาน จัดกลุ่มประเด็นปัญหาและข้อคิดเห็นในการปริวรรตเอกสารโบราณและการปริวรรตคำราชาโบราณอีสาน นำเสนอข้อมูลโดยการแจกแจงความถี่ (Frequencies) และคำนวณหาค่าสัดส่วนหรือร้อยละ (Percentage)

3.1.2 การศึกษาโครงสร้างการบันทึกคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ที่ส่งผลต่อประสิทธิภาพในการปริวรรต

3.1.2.1 วิธีการวิจัย ใช้วิธีวิจัยเชิงคุณภาพ เก็บข้อมูลโดยการศึกษาคำราชาอีสานที่ปริวรรตจากคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย

3.1.2.2 กลุ่มตัวอย่าง ได้แก่ คำราชาอีสานที่ปริวรรตจากคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย โดยเป็นเล่มที่สามารถติดต่อสอบถามข้อมูลเพิ่มเติมจากผู้ปริวรรตได้ ในกรณีที่เกิดข้อสงสัย จำนวนทั้งหมด 7 เล่ม ได้แก่ คำราชาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 1 - 5 และคำราชาวัดท่าม่วง อำเภอเสลภูมิ จังหวัดร้อยเอ็ด เล่ม 1 - 2 ทั้งหมดปริวรรตโดยนักภาษาโบราณในสังกัด สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม และเลือกศึกษาเฉพาะผู้ที่ปริวรรตมาจากอักษรไทยน้อย จำนวนทั้งสิ้น 11 ผูก 459 หน้าลาน

3.1.2.3 เครื่องมือที่ใช้ในการรวบรวมข้อมูล คือ แบบวิเคราะห์โครงสร้างการบันทึกตำรายาโบราณอีสาน โดยประเด็นการศึกษาประกอบด้วย 2 ส่วน ได้แก่ (1) เนื้อหาที่บันทึกในตำรายาและตำรับยาอีสานโบราณ และ (2) โครงสร้างการเขียนตำรายาและตำรับยาอีสานโบราณ

3.1.2.4 การวิเคราะห์ข้อมูล ใช้วิธีวิเคราะห์เนื้อหาและลักษณะโครงสร้างการเขียนตำรายาและตำรับยาอีสานโบราณ จัดกลุ่มเนื้อหาและหารูปแบบโครงสร้างการเขียนตำรายาและตำรับยาอีสานโบราณ

3.1.3 การศึกษาลักษณะของอักษรไทยน้อย และอักษรวิธีของอักษรไทยน้อยที่ใช้บันทึกตำรายาโบราณอีสานในโบราณ

การศึกษานี้ เป็นการศึกษาเพื่อยืนยันลักษณะของอักษรไทยน้อยที่เป็นปัญหาต่อการปริวรรตในขั้นตอนการทำคำปริวรรตอักษรไทยปัจจุบัน (ถอดอักษรไทยน้อยเป็นอักษรไทยปัจจุบัน) และการทำคำอ่านภาษาไทยถิ่นอีสาน กระบวนการวิจัยมีดังนี้

3.1.3.1 วิธีการวิจัย ใช้วิธีวิจัยเชิงคุณภาพ โดยการวิเคราะห์ลักษณะอักษรและภาษาที่ใช้บันทึกตำรายาโบราณอีสาน จากสำเนาภาพถ่ายดิจิทัลของตำรายาโบราณอีสานที่บันทึกในโบราณด้วยอักษรไทยน้อย

3.1.3.2 กลุ่มตัวอย่าง ได้แก่ สำเนาภาพถ่ายดิจิทัลตำรายาโบราณอีสานที่บันทึกในโบราณด้วยอักษรไทยน้อยที่ทำคำอ่านภาษาไทยถิ่นอีสานเอาไว้แล้ว และสามารถสอบถามข้อมูลจากผู้ปริวรรตได้ ได้แก่ สำเนาภาพถ่ายดิจิทัลตำรายาวิคมาหัชย จังหวัดมหาสารคาม เล่ม 1 - 5 และภาพถ่ายดิจิทัลตำรายาวิคทำม่วง จังหวัดร้อยเอ็ด เล่ม 1 - 2 จากสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม โดยคัดเลือกนำมาศึกษาเฉพาะสำเนาภาพถ่ายดิจิทัลของตำรายาโบราณอีสานที่บันทึกในโบราณด้วยอักษรไทยน้อย จำนวน 11 ผูก 459 ภาพ (เป็นภาพตำรายาโบราณอีสานต้นฉบับของตำรายาอีสานซึ่งเป็นกลุ่มตัวอย่างในหัวข้อ 3.1.2)

3.1.3.3 เครื่องมือที่ใช้ในการรวบรวมข้อมูล คือ แบบวิเคราะห์ลักษณะอักษรไทยน้อยที่ใช้บันทึกตำรายาโบราณอีสานในโบราณ ประกอบด้วยรายการสำหรับรวบรวมข้อมูล 2 ด้าน ได้แก่ ลักษณะทั่วไป และอักษรวิธีหรือวิธีการสะกดคำของอักษรไทยน้อย ที่ใช้บันทึกตำรายาโบราณอีสานที่เป็นปัญหาในการปริวรรต ในขั้นตอนการทำคำปริวรรตอักษรไทยปัจจุบันและคำอ่านภาษาไทยถิ่นอีสาน

3.1.3.4 การวิเคราะห์ข้อมูล ใช้การวิเคราะห์ข้อมูลเชิงคุณภาพ โดยหาลักษณะการสะกดคำในตำรายาอีสานด้วยอักษรไทยน้อย ความแตกต่างระหว่างการสะกดคำด้วยอักษรไทยน้อยและคำที่สะกดด้วยอักษรไทยปัจจุบัน

3.2 ระยะที่ 2 การออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

การวิจัยระยะที่ 2 เป็นการออกแบบและพัฒนากรอบการทำงานเพื่อนำมาใช้แก้ปัญหาในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน โดยนำผลการศึกษาระยะที่ 1 ได้แก่ (1) การศึกษากระบวนการปริวรรตเอกสารโบราณและปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยในใบลาน (2) การศึกษาโครงสร้างการบันทึกคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ที่ส่งผลต่อประสิทธิภาพในการปริวรรต และ (3) การศึกษาลักษณะของอักษรไทยน้อย และอักษรวิธีของอักษรไทยน้อยที่ใช้บันทึกคำราชาโบราณอีสานในใบลานมาเป็นข้อมูลพื้นฐานสำหรับการออกแบบและพัฒนากรอบการทำงาน

วิธีการวิจัยแบ่งออกเป็น 4 ส่วน ได้แก่ (1) การออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (2) การพัฒนาออนโทโลยีคำราชาโบราณอีสาน (3) การออกแบบขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน และ (4) การสร้างข้อมูลชุดทดสอบประสิทธิภาพด้านความถูกต้องของคลังศัพท์ ออนโทโลยี และขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน รายละเอียดวิธีการวิจัยในแต่ละส่วนมีดังนี้

3.2.1 การออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

การวิจัยนี้ เป็นการวิจัยเพื่อสร้างฐานความรู้สำหรับใช้ในกระบวนการตัดคำ และแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานแบบไม่พึ่งบริบทและใช้คลังศัพท์ (Lexicon) เป็นฐานในการทำคำอ่าน การออกแบบและพัฒนาคลังศัพท์ที่ใช้วิธีตามขั้นตอนการออกแบบและพัฒนาฐานข้อมูล (ศิริลักษณ์ โรจกิจอำนวย, 2552) และหลักการพัฒนาคลังศัพท์แบบใช้คลังคำ (Corpus) เป็นฐาน (วิโรจน์ อรุณมานะกุล, 2553) รายละเอียดมีดังนี้

3.2.1.1 วิเคราะห์ลักษณะการอธิบายคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสานที่มีอยู่ในปัจจุบัน และความเหมาะสมในการนำมาใช้แปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน โดยวิเคราะห์จากลักษณะโครงสร้างการอธิบายคำศัพท์ทั้งจากคลังศัพท์ภาษาไทยมาตรฐานและคลังศัพท์ภาษาไทยถิ่นอีสาน รวมทั้งสิ้น 8 เล่ม ได้แก่ (1) สารานุกรมภาษาไทยถิ่นอีสาน-ไทย-อังกฤษ (ปรีชา พิณทอง, 2532) (2) ประมวลคำศัพท์ภาษาลาวในเอกสารโบราณ เล่ม 1 - 5 (หอสมุดแห่งชาติ, 2545) (3) พจนานุกรม 3 ภาษา (ไทยน้อย-อังกฤษ-ไทย) (รัตนจันทร์เทาว์, 2558) (4) พจนานุกรมภาคอีสาน-ภาคกลาง ฉบับปณิธานสมเด็จพระมหาวิรวงศ์

(ดิศุสมหาเถระ) (สมเด็จพระมหาธีรราชเจ้า (ดิศุสมหาเถระ), 2515) (5) พจนานุกรมภาษาถิ่นอีสาน: เว้าอีสาน (บุญเกิด พิมพ์วรรณา และ นภาพร พิมพ์วรรณา, 2545) (6) พจนานุกรมภาษาไทยถิ่นอีสาน ฉบับ คำพูน บุญทวี (วีระ สดสังข์ และ ลันนา เจริญสิทธิชัย, 2548) (7) พจนานุกรมภาษาไทยถิ่นอีสาน-ไทยกลาง (สำลี รักสุทธี, 2554) และ (8) พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (ราชบัณฑิตยสถาน, 2557)

3.2.1.2 การออกแบบและพัฒนาคำศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน แบ่งออกเป็น 3 ขั้นตอนหลัก ดังนี้

1) ออกแบบและสร้างฐานข้อมูลคำศัพท์ มีขั้นตอน ดังนี้

1.1) กำหนดลักษณะคำศัพท์สำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

1.2) ออกแบบโครงสร้างฐานข้อมูลคำศัพท์ ระดับแนวคิด (Conceptual Schema Design) วิธีดำเนินการ ดังนี้ นำผลการกำหนดลักษณะคำศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน และผลการวิเคราะห์ลักษณะการอธิบายโครงสร้างคำจากข้อ 3.2.1.1 มาใช้เป็นแนวทางในการออกแบบโครงสร้างฐานข้อมูลคำศัพท์ ในระดับแนวคิด นำเสนอผลการออกแบบด้วยแผนภาพแสดงความสัมพันธ์ระหว่างเอนทิตี (Entity Relationship Diagram) และพจนานุกรมข้อมูล (Data Dictionary)

1.3) ออกแบบฐานข้อมูลคำศัพท์ ระดับกายภาพ (Physical Design) วิธีดำเนินการ ดังนี้ (1) กำหนดเครื่องมือสำหรับสร้างและจัดการฐานข้อมูล ได้แก่ ระบบปฏิบัติการ ไมโครซอฟต์วินโดวส์ 10 และ โปรแกรมจัดการฐานข้อมูล SQLite Studio (2) อุปกรณ์จัดเก็บข้อมูล ได้แก่ เครื่องคอมพิวเตอร์แบบพกพา ที่มีหน่วยประมวลผลกลาง Intel Core i7 มีหน่วยความจำหลัก (RAM) ขนาด 16 กิกะไบต์ และมีหน่วยความจำสำรอง (Hard Disk) ขนาด 540 กิกะไบต์ และ กำหนดตำแหน่งการจัดเก็บฐานข้อมูล คือ หน่วยความจำสำรองของอุปกรณ์จัดเก็บข้อมูลที่กำหนดในงานวิจัย (3) กำหนดกลุ่มผู้มีสิทธิใช้งานฐานข้อมูล แบ่งเป็น 3 กลุ่ม ได้แก่ ผู้ดูแลฐานข้อมูล โปรแกรมเมอร์ และผู้ใช้ทั่วไป

1.4) สร้างฐานข้อมูลคำศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน มีวิธีดำเนินการ ดังนี้ (1) ติดตั้งโปรแกรมจัดการฐานข้อมูล SQLite Studio (2) สร้างฐานข้อมูลคำศัพท์ตามโครงสร้างฐานข้อมูลที่ออกแบบไว้ในขั้นตอนที่ 1.2)

2) วิเคราะห์เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสานที่ใช้ในคลังศัพท์ภาษาไทยถิ่นอีสาน (เอกสารชุดเดียวกันกับหัวข้อ 3.2.1.2) เพื่อนำมาใช้เป็นเกณฑ์สะกดคำศัพท์ภาษาไทยถิ่นอีสานในคลังคำที่จะพัฒนาขึ้นมาในงานวิจัยนี้ โดยใช้วิธีเลือกเกณฑ์สะกดคำที่มีความสอดคล้องกับลักษณะของการออกเสียงในภาษาไทยถิ่นอีสาน (วิไลศักดิ์ กิ่งคำ, 2556) และหน่วยเสียงพยัญชนะและสระตามเอกสารโบราณ

3) การวิเคราะห์และกำหนดชนิดคำศัพท์ (Part-Of-Speech: POS) สำหรับกำกับชนิดของคำในคลังศัพท์ โดยใช้วิธีวิเคราะห์กฎทางวากยสัมพันธ์ของภาษาไทย มาตรฐาน (Syntactic Roles) และผลการวิเคราะห์โครงสร้างการเขียนคำราชาโบราณอีสานมาเป็นเกณฑ์การแบ่งประเภทชนิดของคำศัพท์

4) การรวบรวมคำศัพท์และจัดเก็บคำศัพท์ในฐานข้อมูล คือขั้นตอนการรวบรวมคำศัพท์และนำไปบันทึกจัดเก็บในฐานข้อมูลคลังศัพท์ที่สร้างเอาไว้ มีขั้นตอนดำเนินการดังนี้

4.1) กำหนดโปรแกรมสำหรับใช้เตรียมคำศัพท์ ได้แก่ โปรแกรม ไมโครซอฟต์เวิร์ดใช้สำหรับแบ่งแยกประโยคในหนังสือคำราชาให้เป็นคำ โปรแกรม ไมโครซอฟต์ เอ็กเซลใช้สำหรับจัดเรียงคำ ตัดคำซ้ำ และจัดเตรียมไฟล์คำศัพท์สำหรับนำเข้าสู่ระบบฐานข้อมูลคลังศัพท์

4.2) รวบรวมคำศัพท์ มีขั้นตอนดังนี้

4.2.1) รวบรวมข้อมูลคำศัพท์จากคำราชาอีสานที่ปริวรรตมาจากคำราชาโบราณอีสานที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย ได้แก่ คำราชาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 5 และคำราชาวัดท่าม่วง จังหวัดร้อยเอ็ด เล่ม 2 โดยขออนุญาตใช้ข้อมูลและทำสำเนาไฟล์ข้อมูลแบบไมโครซอฟต์เวิร์ดของหนังสือคำราชาอีสานทั้ง 2 เล่ม จากสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม คัดเลือกเฉพาะผู้ที่ปริวรรตมาจากอักษรไทยน้อย จำนวนทั้งสิ้น 11 ชุด 459 หน้าลาน

4.2.2) ตัดคำ โดยนำข้อความในไฟล์หนังสือคำราชาอีสานมาตัดคำด้วยมือ ด้วยโปรแกรมไมโครซอฟต์เวิร์ด โดยใช้เกณฑ์การตัดคำที่ยาวที่สุดในรอบแรก และตัดคำแบบสั้นที่สุดสำหรับคำประสมในรอบที่ 2 เพื่อแยกคำประสมเป็นคำโดดเพิ่มเติม (คำประสมคำเดิมยังคงมีอยู่)

4.2.3) กำกับชนิดของคำ โดยใช้ชนิดของคำตามผลการวิจัยการวิเคราะห์และกำหนดชนิดคำศัพท์ (Part-Of-Speech: POS) สำหรับกำกับชนิดของคำในคลังศัพท์ มาใช้กำกับชนิดของคำ

4.2.4) ตรวจสอบการสะกดคำศัพท์ให้เป็นมาตรฐานเดียวกัน โดยใช้เกณฑ์จากขั้นตอนการวิเคราะห์เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสานที่ใช้ในคลังศัพท์ภาษาไทยถิ่นอีสาน มาเป็นเกณฑ์ตรวจสอบและแก้ไขการสะกดคำให้เป็นมาตรฐานเดียวกัน

4.2.5) ทำคำปริวรรตโดยการนำคำอ่านมาแปลงกลับเป็นคำปริวรรต โดยใช้เกณฑ์การปริวรรตอักษรไทยน้อยเป็นอักษรไทยมาเป็นเกณฑ์ในการแปลง

4.2.6) ทำคำปรัวรรตให้กับคำอ่านแต่ละคำ โดยใช้หลักการปรัวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบันมาเป็นเกณฑ์ในการทำคำปรัวรรต แล้วตรวจความถูกต้องโดยเทียบกับคำศัพท์อักษรไทยน้อยในสำเนาภาพดิจิทัลตำรายาโบราณอีสานต้นฉบับ

4.2.7) ตัดคำซ้ำ โดยคัดลอกรายการคำศัพท์ที่ตัดคำเรียบร้อยแล้วด้วยโปรแกรมไมโครซอฟต์เวิร์ดไปบันทึกในโปรแกรมไมโครซอฟต์เอ็กเซล จัดเรียงคำศัพท์ตามตัวอักษร ก-ฮ และตัดคำที่เขียนสะกดซ้ำกันออกไป

4.2.8) กำหนดรายละเอียดของคำศัพท์ตามโครงสร้างของฐานข้อมูลคลังศัพท์ คำปรัวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน โดยสืบค้นข้อมูลจากแหล่งข้อมูลต่าง ๆ ได้แก่ พจนานุกรมภาษาไทยถิ่นอีสาน จำนวน 6 เล่ม (เอกสารชุดเดียวกันกับหัวข้อ 3.2.1.2) สารานุกรมรวมหลักเกณฑ์กรมไทย (วุฒิ วุฒิชรรมเวช, 2540) เว็บไซต์ฐานข้อมูลสมุนไพร คณะเภสัชศาสตร์ มหาวิทยาลัยอุบลราชธานี (www.phargarden.com) และเว็บไซต์เมดไทย (www.medthai.com) โดยคำศัพท์ที่ไม่สามารถค้นหารายละเอียดจากแหล่งข้อมูลดังกล่าวได้จะใช้วิธีสอบถามจากนักภาษาโบราณ หรือหมอยาอีสาน และระบุแหล่งสืบค้นข้อมูลเป็น “สัมภาษณ์”

4.2.9) ตรวจสอบความถูกต้องของคำปรัวรรต คำอ่าน และรายละเอียดอื่น ๆ โดยนักภาษาโบราณซึ่งเป็นผู้ปรัวรรตตำรายาฉบับที่นำมาศึกษา

4.2.10) แบ่งแยกข้อมูลคำศัพท์ตามโครงสร้างฐานข้อมูลที่ออกแบบไว้ บันทึกในแบบไฟล์ CSV

5) บันทึกคำศัพท์และรายละเอียดข้อมูลคำศัพท์ลงในฐานข้อมูล คลังศัพท์ โดยนำคำศัพท์ที่ถูกจัดเก็บในรูปแบบไฟล์ CSV นำเข้าสู่ฐานข้อมูลด้วยคำสั่ง โปรแกรมจัดการฐานข้อมูล SQLite Studio

6) ประเมินประสิทธิภาพคลังศัพท์ แบ่งเป็น 2 ด้าน ได้แก่ การประเมินประสิทธิภาพเชิงคุณภาพด้านความเหมาะสมในการออกแบบและพัฒนาคลังศัพท์ และการประเมินประสิทธิภาพเชิงปริมาณ คือ การประเมินความถูกต้องในการแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน โดยการประเมินผลจะดำเนินการในการวิจัยระยะที่ 3 คือ การประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

3.2.2 การพัฒนาออนโทโลยีตำรายาโบราณอีสาน

การวิจัยนี้ เป็นการวิจัยเพื่อสร้างเครื่องมือสำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ที่คำปริวรรตต้องพึงบริบทในการทำคำอ่านให้ถูกต้อง (คำปริวรรตประเภท 1:M และ M:N) ขั้นตอนการวิจัยกำหนดตามตัวแบบการพัฒนาออนโทโลยีตามแนวคิดของ นอย และแม็กกวินเนส (Noy, N.F., and McGuinness, D.L., 2001) แบ่งเป็น 8 ขั้นตอน รายละเอียดมีดังนี้

3.2.2.1 กำหนดขอบเขตความรู้และขอบเขตของออนโทโลยีตำรายาโบราณอีสาน โดยใช้ผลการศึกษาปัญหาในระยะที่ 1 มากำหนด

3.2.2.2 พิจารณานำออนโทโลยีที่มีอยู่แล้วนำกลับมาใช้ใหม่ โดยนำออนโทโลยีด้านยาสมุนไพรที่บันทึกไว้ในเอกสารโบราณ (นิสาชล จำนวนศรี, 2562) มาปรับปรุงให้ครอบคลุมความรู้ในตำรายาโบราณอีสานและความต้องการฐานความรู้สำหรับการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

3.2.2.3 กำหนดรายการคำศัพท์ (Terms) ที่จะกำหนดให้มีในออนโทโลยี โดยนำคำศัพท์ที่ใช้เขียนในตำรายาโบราณอีสานมากำหนดเป็นรายการคำศัพท์

3.2.2.4 กำหนดคลาสและลำดับชั้น (Define the Classes and Class Hierarchy) ได้แก่ นำคำศัพท์มาจัดกลุ่มเนื้อหาเพื่อกำหนดเป็นชื่อคลาส แล้วนำคลาสมาจัดลำดับชั้นโดยใช้วิธีแบบบนลงล่าง (Top-down) (Uschold, M., and Gruninger, M., 1996)

3.2.2.5 กำหนดคุณสมบัติ (Properties) ของคลาส หรือสล็อต (Slots) สำหรับอธิบายคลาสแต่ละคลาส

3.2.2.6 กำหนดชนิดข้อมูล (Data Type) ของแต่ละแอททริบิวต์ เช่น ชนิดข้อมูลแบบสายอักขระ (String) ตัวเลข (Number) ตรรกศาสตร์ (Boolean) เป็นต้น

3.2.2.7 สร้างคำศัพท์ในแต่ละคลาส ได้แก่ การนำคำศัพท์ตำรายาโบราณอีสานไปสร้างในออนโทโลยีด้วยโปรแกรมโปรทีเจ รุ่น 5.5 (Protégé 5.5)

3.2.2.8 ประเมินผลออนโทโลยี ใช้วิธีป้อนคำอ่านภาษาไทยถิ่นอีสานที่ต้องการสอบถามพร้อมคำบริบทซ้ายและคำบริบทขวา ไปค้นหาในออนโทโลยีว่าคำอ่านและบริบทมีความสอดคล้องกันหรือไม่ นำผลลัพธ์คำอ่านที่ได้ไปตรวจความถูกต้องโดยการเทียบกับเฉลยของนักภาษาโบราณ

3.2.3 การออกแบบขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

เป็นการออกแบบขั้นตอนวิธีเพื่อใช้แปลงคำรับยาโบราณอีสานที่เขียนสะกดคำแบบคำปริวรรตของอักษรไทยน้อยไปเป็นคำอ่านภาษาไทยถิ่นอีสาน รายละเอียดวิธีวิจัยมีดังนี้

3.2.3.1 ทบทวนวรรณกรรมด้านการประมวลผลภาษาธรรมชาติ การแปลภาษาด้วยเครื่อง การปริวรรตเอกสารโบราณแบบอัตโนมัติ กระบวนการปริวรรตเอกสารโบราณ และกระบวนการปริวรรตคำราชาโบราณอีสาน

3.2.3.2 วิเคราะห์กระบวนการประมวลผลภาษาธรรมชาติ การแปลภาษาด้วยเครื่อง กระบวนการปริวรรตเอกสารโบราณแบบอัตโนมัติ กระบวนการปริวรรตเอกสารโบราณ และกระบวนการปริวรรตคำราชาโบราณอีสาน สรุปผลการวิเคราะห์เป็นแผนภาพแสดงกระบวนการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

3.2.3.3 ออกแบบขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน โดยใช้ผลการวิเคราะห์ในข้อที่ 3.2.3.2 นำเสนอขั้นตอนวิธีด้วยผังงาน (Flowchart)

3.2.4 การสร้างข้อมูลชุดทดสอบประสิทธิภาพการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

การเตรียมข้อมูลสำหรับการประเมินประสิทธิภาพ คือการสร้างชุดข้อมูล (คำรับยาโบราณอีสานในแบบคำปริวรรตของอักษรไทยน้อย) สำหรับการประเมินความถูกต้องในการแปลงคำปริวรรตเป็นคำอ่าน โดยรวบรวมข้อมูลจากตำรายาวิกรมหาชัย จังหวัดมหาสารคาม เล่มที่ 1 - 4 และหนังสือตำรายาวิกรมท่าม่วง อำเภอเสลภูมิ จังหวัดร้อยเอ็ด เล่ม 1 ซึ่งเป็นข้อมูลคนละชุดกับที่นำมาใช้สร้างคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน และออนไลน์คำราชาโบราณอีสาน นำมาแบ่งแยกเป็นคำรับยาและจัดเก็บในรูปแบบไฟล์ข้อความ (Text file) ขั้นตอนดำเนินการมีดังนี้

3.2.4.1 นำคำราชาโบราณอีสานที่รวบรวมได้มาตัดแยกเป็นคำรับยาแต่ละคำรับ

3.2.4.3 แปลงคำรับยา (สะกดคำแบบปัจจุบัน) ย้อนกลับไปเป็นคำรับยาแบบคำปริวรรต (สะกดคำแบบโบราณ)

3.2.4.4 ตรวจสอบความถูกต้องของคำปริวรรตกับอักษรไทยน้อยในเอกสารต้นฉบับ

3.2.4.5 บันทึกจัดเก็บคำรับยาที่แปลงเป็นคำปริวรรตเรียบร้อยแล้วในรูปแบบไฟล์ข้อความ (Text File)

3.3 ระยะที่ 3 การประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

3.3.1 การประเมินประสิทธิภาพเชิงปริมาณกรอบการทำงานในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

เป็นการประเมินความถูกต้องของการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ขั้นตอนและวิธีการประเมินผลมีดังนี้

3.3.1.1 พัฒนาโปรแกรมแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ด้วยโปรแกรมภาษาไพธอน 3.8 (Python 3.8) ตามกรอบการทำงานและขั้นตอนวิธีที่พัฒนาขึ้นมาในงานวิจัยนี้

3.3.1.2 นำไฟล์ข้อมูลชุดทดสอบที่เตรียมไว้สำหรับจะประเมินประสิทธิภาพความเข้าสู่โปรแกรมที่พัฒนาขึ้นในข้อ 3.2.4 เพื่อประมวลผลเป็นคำอ่านภาษาไทยถิ่นอีสาน

3.3.1.3 นำผลลัพธ์คำอ่านภาษาไทยถิ่นอีสานที่ได้จากการประมวลผลด้วยโปรแกรมที่พัฒนาขึ้นไปตรวจสอบความถูกต้อง โดยเปรียบเทียบกับผลลัพธ์คำเฉลยของนักภาษาโบราณและนับจำนวนคำอ่านภาษาไทยถิ่นอีสานที่ถูกต้องของคำรับยาแต่ละคำรับ

3.3.1.4 คำนวณหาระดับความถูกต้องของคำอ่านภาษาไทยถิ่นอีสานของคำรับยาแต่ละคำรับ โดยหาค่าร้อยละความถูกต้อง (Accuracy) ของคำอ่านภาษาไทยถิ่นอีสาน ด้วยสมการทางคณิตศาสตร์ ดังนี้

$$\text{ร้อยละความถูกต้องของคำอ่านภาษาไทยถิ่นอีสาน} = \frac{\text{จำนวนคำอ่านภาษาไทยถิ่นอีสานที่ถูกต้อง}}{\text{จำนวนคำอ่านภาษาไทยถิ่นอีสานทั้งหมด}} \times 100$$

3.3.2 การประเมินประสิทธิภาพเชิงคุณภาพกรอบการทำงานในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

เป็นการประเมินคุณภาพด้านความเหมาะสมของกรอบการทำงานในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานในด้านความถูกต้องของผลลัพธ์ ความเร็วในการประมวลผล ความครอบคลุมและเพียงพอของคลังคำตำรายาอีสานโบราณ ความเหมาะสมในการนำมาใช้ปวิวรรตตำรายาอีสานโบราณ และประโยชน์ที่ได้รับสำหรับการปวิวรรตตำรายาอีสานโบราณ โดยสอบถามความคิดเห็นของกลุ่มตัวอย่างซึ่งเป็นผู้ใช้ที่เป็นผู้เชี่ยวชาญอักษรและภาษาโบราณ และผู้ใช้ประโยชน์จากตำรายาอีสานโบราณ วิธีประเมินมีดังนี้

3.3.2.1 กลุ่มตัวอย่างที่ใช้ในการวิจัย แบ่งเป็น 2 กลุ่ม โดยกลุ่มที่ 1 เป็นกลุ่มผู้เชี่ยวชาญด้านอักษรและภาษาโบราณ และกลุ่มที่ 2 เป็นกลุ่มผู้ใช้ประโยชน์จากเนื้อหาที่บันทึกในตำราภาษาโบราณอีสาน การกำหนดกลุ่มตัวอย่างเป็น 2 กลุ่ม มีวัตถุประสงค์เพื่อให้ทราบความคิดเห็นของทั้ง 2 กลุ่ม ซึ่งกลุ่มเป็นผู้เชี่ยวชาญและผู้ที่ไม่เชี่ยวชาญด้านอักษรและภาษาโบราณ รายละเอียดมีดังนี้

1) กลุ่มที่ 1 กลุ่มผู้เชี่ยวชาญ ได้แก่ นักภาษาโบราณที่อ่านอักษรไทยน้อยได้ รู้และเข้าใจภาษาไทยถิ่นอีสาน สังกัดประจำอยู่ในหน่วยงานซึ่งมีหน้าที่อนุรักษ์ สืบค้น และเผยแพร่ความรู้ในเอกสารโบราณโดยตรง จาก 2 หน่วยงาน ได้แก่ กลุ่มหนังสือตัวเขียนและจารึกหอสมุดแห่งชาติ ท่าवासกรี กรุงเทพมหานคร และหน่วยงานในส่วนภูมิภาค จำนวน 7 คน และสถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม จำนวน 5 คน จำนวนทั้งสิ้น 12 คน

2) กลุ่มที่ 2 ผู้ใช้ตำราภาษาโบราณอีสานเพื่อการศึกษา ค้นคว้า ได้แก่ นักศึกษา ระดับปริญญาตรี ที่กำลังศึกษาในหลักสูตรแพทยแผนไทยหรือสาขาที่เกี่ยวข้อง ที่ต้องค้นคว้าหาความรู้จากตำราภาษาโบราณอีสาน และสามารถอ่านอักษรไทยน้อยได้ ได้แก่ นักศึกษาระดับปริญญาตรี ชั้นปีที่ 4 หลักสูตรแพทยแผนไทยประยุกต์ คณะแพทยศาสตร์ มหาวิทยาลัยมหาสารคาม ที่ผ่านการเรียนรายวิชา ตำราภาษาโบราณในเอกสารโบราณ แล้ว จำนวนทั้งสิ้น 27 คน

3.3.2.2 การรวบรวมข้อมูลและเครื่องมือรวบรวมข้อมูล มีรายละเอียดดังนี้

1) เครื่องมือและการสร้างเครื่องมือรวบรวมข้อมูล ประกอบด้วย ระบบต้นแบบสำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน และแบบประเมินประสิทธิภาพประกอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน รายละเอียดการสร้างเครื่องมือมี ดังนี้

1.1) ระบบต้นแบบสำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน สร้างตามตัวแบบวงจรการพัฒนาระบบ (System Development Life Cycle: SDLC)

1.2) แบบประเมินประสิทธิภาพประกอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน มีรายละเอียดการสร้าง ได้แก่

1.2.1) ศึกษาทฤษฎีการสร้างแบบประเมินประสิทธิภาพระบบงานคอมพิวเตอร์จากตำราและเอกสารที่เกี่ยวข้องเพื่อนำมาใช้เป็นแนวทางในการสร้างแบบประเมิน

1.2.2) กำหนดหัวข้อและสร้างแบบประเมิน โดยแบ่งแบบประเมินออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 ข้อมูลทั่วไปเกี่ยวกับผู้ตอบแบบประเมิน ลักษณะแบบประเมินเป็นแบบปลายเปิด จำนวน 4 ข้อ ได้แก่ ประสิทธิภาพในการทำงานด้านการปริวรรต

อักษรไทยโบราณ อักษรและภาษาโบราณที่เชี่ยวชาญมากที่สุด ประสบการณ์ในการปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรไทยน้อย และประสบการณ์ในการปริวรรตตำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อย ตอนที่ 2 ประสิทธิภาพการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ลักษณะแบบประเมินเป็นแบบปลายปิด ประกอบด้วยประเด็นคำถาม 3 ด้าน ได้แก่ ประสิทธิภาพด้านความถูกต้องของผลลัพธ์ ความเร็วในการประมวลผล ประสิทธิภาพด้านความครอบคลุมและเพียงพอของคลังศัพท์ตำรายาอีสานโบราณ ความเหมาะสมในการนำมาใช้ปริวรรตตำรายาอีสานโบราณ และประโยชน์ที่ได้รับสำหรับการปริวรรตตำรายาอีสานโบราณ

1.2.3) เกณฑ์การให้คะแนนการประเมินประสิทธิภาพ แบ่งเป็น 5 ระดับ (Rating Scale) ตามวิธีของลิเคอร์ท (Likert, 1961) ได้แก่ มากที่สุด มาก ปานกลาง น้อย และน้อยที่สุด โดยมีมาตรส่วนการให้คะแนน ดังนี้

5	หมายถึง	มากที่สุด
4	หมายถึง	มาก
3	หมายถึง	ปานกลาง
2	หมายถึง	น้อย
1	หมายถึง	น้อยที่สุด

1.2.4) การหาประสิทธิภาพของแบบประเมิน มีขั้นตอนการดำเนินการ ได้แก่ นำแบบประเมินที่สร้างขึ้นไปให้ผู้เชี่ยวชาญด้านการออกแบบและพัฒนาระบบงานทางคอมพิวเตอร์ จำนวน 5 คน ทำการประเมินความเหมาะสมและความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ของการประเมิน โดยใช้วิธีการประเมินค่าดัชนีความสอดคล้อง (Item of item Objective Congruence Index: IOC) ที่มีคะแนน IOC ตั้งแต่ 0.75 ขึ้น นำแบบประเมินที่ผ่านการประเมินแล้วมาปรับปรุงแก้ไขเพื่อนำไปใช้กับกลุ่มตัวอย่างต่อไป

1.2.5) การวิเคราะห์ข้อมูลประสิทธิภาพ โดยหาค่าเฉลี่ยคะแนนประสิทธิภาพของกลุ่มตัวอย่าง (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) (สุทธิ ชัดติยะ และ วิไลลักษณ์ สุวจิตตานน, 2554) ดังนี้

$$\bar{X} = \frac{\sum X_i}{n}$$

เมื่อ \bar{X} หมายถึง ค่าเฉลี่ยคะแนนประสิทธิภาพ

$\sum X_i$ หมายถึง ผลรวมคะแนนประสิทธิภาพ

n หมายถึง จำนวนกลุ่มตัวอย่าง

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

เมื่อ SD หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

$\sum(x_i - \bar{x})^2$ หมายถึง ผลรวมของผลต่างกำลังสองของ
คะแนนแต่ละข้อคำถามและค่าเฉลี่ย
n หมายถึง จำนวนกลุ่มตัวอย่าง

ดังนี้

การคำนวณความกว้างของอัตรภาคชั้นของค่าเฉลี่ย ใช้สมการทางคณิตศาสตร์

$$\begin{aligned} \text{ความกว้างของอัตรภาคชั้น} &= \frac{\text{คะแนนสูงสุด} - \text{คะแนนต่ำสุด}}{\text{จำนวนชั้น}} \\ &= \frac{5-1}{5} \\ &= 0.80 \end{aligned}$$

เกณฑ์การแปลความหมายคะแนนเฉลี่ย ใช้ช่วงคะแนนเฉลี่ย ดังนี้

4.21-5.00	หมายถึง	ดีมาก
3.81-4.20	หมายถึง	ดี
2.21-3.00	หมายถึง	ปานกลาง
1.81-2.20	หมายถึง	น้อย
1.00-1.80	หมายถึง	น้อยมาก

บทที่ 4

ผลการศึกษาปัญหาการปริวรรตตำรายาโบราณอีสาน ที่บันทึกในใบลานด้วยอักษรไทยน้อย

การนำเสนอผลการวิจัยเรื่อง การใช้ออนโทโลยีเป็นฐานในการแปลงคำปริวรรตของ อักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน นำเสนอผลการวิจัยแบ่งเป็น 2 ส่วน ได้แก่ ผลการศึกษาปัญหาการปริวรรตตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย (บทที่ 4) และผลการพัฒนาและประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของ อักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน (บทที่ 5) โดยในบทนี้จะเป็นการนำเสนอผล การศึกษาส่วนที่ 1 มีดังนี้

4.1 การศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตตำรายา โบราณอีสานที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย

4.1.1 ข้อมูลส่วนตัว ประสบการณ์ และความเชี่ยวชาญด้านการปริวรรตเอกสาร โบราณ

4.1.2 ปัญหาการปริวรรตเอกสารโบราณ

4.1.3 กระบวนการปริวรรตเอกสารโบราณ

4.1.4 ปัญหาการปริวรรตตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษร ไทยน้อย

4.2 การศึกษาโครงสร้างการเขียนตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษร ไทยน้อยที่ส่งผลต่อประสิทธิภาพในการปริวรรต

- 4.2.1 องค์ประกอบด้านเนื้อหาที่บันทึกในตำรับยาอีสานโบราณ
- 4.2.2 โครงสร้างการเขียนตำรายาโบราณอีสาน
- 4.2.3 การจัดกลุ่มคำศัพท์ที่ใช้เขียนในตำรายาโบราณอีสาน

4.3 การศึกษาลักษณะอักษรไทยน้อย และอักษรวิธีอักษรไทยน้อยที่ใช้บันทึกตำรายา โบราณอีสานในใบลาน

- 4.3.1 รูปร่างอักษรไทยน้อย
- 4.3.2 ภาษา
- 4.3.3 อักษรวิธีการประสมอักษรไทยน้อยเป็นคำ

4.3.4 ลักษณะการสะกดคำของอักษรไทยน้อย

4.4 ลักษณะอักษรไทยน้อยที่ส่งผลกระทบต่อการศึกษาโบราณอีสานแบบอัตโนมัติ

4.4.1 ปัญหาการทำคำปริวรรต

4.4.2 ปัญหาการทำคำอ่านภาษาไทยถิ่นอีสาน

4.1 การศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตตำรายาโบราณอีสานที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย

4.1.1 ข้อมูลส่วนตัว ประสบการณ์ และความเชี่ยวชาญด้านการปริวรรตเอกสารโบราณ

การศึกษานี้เก็บรวบรวมข้อมูลจากนักภาษาโบราณ ที่สังกัดใน 2 หน่วยงาน ได้แก่ กลุ่มหนังสือตัวเขียนและจารึก หอสมุดแห่งชาติ ท่าวาสุกรี กรุงเทพมหานคร และงานอนุรักษ์เอกสารโบราณ สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม จำนวนรวมทั้งสิ้น 10 คน เครื่องมือในการรวบรวมข้อมูลคือ แบบสัมภาษณ์แบบกึ่งโครงสร้าง โดยอายุของนักภาษาโบราณอยู่ระหว่าง 29-71 ปี ร้อยละ 50 มีอายุมากกว่า 40 ปี ประสบการณ์ทำงานด้านการปริวรรตเอกสารโบราณ อยู่ระหว่าง 2-35 ปี โดย 9 จาก 10 คนมีประสบการณ์มากกว่า 10 ปี และมีประสบการณ์ 2 ปี 1 คน

วัตถุประสงค์ในการปริวรรตอักษรโบราณของนักภาษาโบราณ ได้แก่ (1) เพื่อเผยแพร่องค์ความรู้ในเอกสารโบราณให้บุคคลทั่วไปนำไปศึกษา หรือนำใช้ประโยชน์ในด้านต่าง ๆ (2) เพื่อศึกษาวิเคราะห์เปรียบเทียบรูปอักษร ภาษา อักษรวิธีอักษรโบราณในยุคต่าง ๆ (3) เพื่ออนุรักษ์ภูมิปัญญาท้องถิ่น และ (4) เพื่อจัดทำทะเบียนเอกสารโบราณ

ประเภทเอกสาร โบราณที่นักภาษาโบราณเคยปริวรรตมากที่สุด 3 อันดับแรก ได้แก่ จารึก สมุดไทย และเอกสารใบลาน เนื้อหาที่เคยปริวรรตมากที่สุด 5 อันดับแรก ได้แก่ ศาสนา วรรณกรรม กฎหมาย ตำรายา และโหราศาสตร์ รายละเอียดดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 หมวดเนื้อหาในเอกสารโบราณที่นักภาษาโบราณเคยปริวรรต

อันดับ	หมวดเนื้อหา	จำนวน (คน)	ร้อยละ
1	ศาสนา	10	100
2	วรรณกรรม	9	90
3	กฎหมาย	8	80
4	ตำรายา	7	70
5	โหราศาสตร์	6	60
6	พิธีกรรม	2	20
7	เวทย์มนต์/คาถา	2	20
8	ดาราศาสตร์	1	10

ความเชี่ยวชาญด้านการปริวรรตเอกสารโบราณ นักภาษาโบราณส่วนใหญ่อ่านอักษรโบราณได้อย่างน้อย 2 อักษร โดยมีหนึ่งอักษรอ่านได้ในระดับมาก และอีกหนึ่งอักษรอ่านได้ในระดับปานกลาง อักษรโบราณที่นักภาษาโบราณมีความเชี่ยวชาญในระดับมาก 3 อันดับแรกคือ อักษรขอมไทย อักษรไทยน้อย และอักษรธรรมอีสาน (อักษรละ 4 คน) รองลงไปคือ อักษรธรรมล้านช้าง (3 คน) และน้อยที่สุดคือ อักษรเทวนาครีและอักษรขอมโบราณ มีเพียง 1 คน อักษรที่นักภาษาโบราณส่วนใหญ่มีความเชี่ยวชาญอ่านได้ในระดับปานกลาง ได้แก่ อักษรไทย อักษรขอมไทย อักษรไทยน้อย อักษรธรรมอีสาน อักษรธรรมล้านช้าง อักษรขอมโบราณ อักษรฝักขาม และอักษรมอญโบราณ รายละเอียดดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 จำนวนนักภาษาโบราณและระดับความเชี่ยวชาญการปริวรรตเอกสารโบราณจำแนกตามอักษรโบราณที่อ่านได้

ลำดับ	ชื่ออักษร	ระดับความเชี่ยวชาญ (คน)				รวม (คน)
		มาก	ปานกลาง	น้อย	ไม่รู้จักรูปอักษร	
1	ขอมไทย	4	6	0	0	10
2	ไทยน้อย	4	6	0	0	10
3	ธรรมอีสาน	4	6	0	0	10
4	ธรรมล้านช้าง	3	5	-	-	8
5	อักษรไทย	2	8	0	0	10
6	ปัลลวะ	2	3	-	1	6
7	ธรรมล้านนา	2	2	-	-	4
8	ขอมโบราณ	1	5	-	-	6

ตารางที่ 4.2 จำนวนนักภาษาโบราณและระดับความเชี่ยวชาญการปริวรรตเอกสารโบราณจำแนกตามอักษรโบราณที่อ่านได้ (ต่อ)

ลำดับ	ชื่ออักษร	ระดับความเชี่ยวชาญ (คน)				รวม (คน)
		มาก	ปานกลาง	น้อย	ไม่รู้จักรูปอักษร	
9	เทวนาครี	1	1	-	4	6
10	ฝักขาม	-	5	-	2	7
11	มอญโบราณ	-	5	-	1	6
12	พม่า	-	2	-	-	2

ความเชี่ยวชาญด้านการอ่านภาษาโบราณพบว่า นักภาษาโบราณส่วนใหญ่อ่านภาษาโบราณได้มากกว่า 1 ภาษา โดยมีความเชี่ยวชาญในระดับปานกลางถึงมาก ภาษาที่เชี่ยวชาญในระดับมาก คือ ภาษาไทยถิ่นอีสาน (5 คน) และ ภาษาบาลี (1 คน) ภาษาที่นักภาษาโบราณส่วนมากอ่านได้ในระดับปานกลาง คือ ภาษาล้านนา (8 คน) รองลงมาคือ ภาษาบาลี (6 คน) และภาษาไทยถิ่นอีสาน (4 คน) ส่วนภาษาที่นักภาษาโบราณส่วนใหญ่อ่านไม่ได้ คือ ภาษามอญ ซึ่งผู้อ่านได้ในระดับปานกลางมีเพียง 1 คน ส่วนอีก 9 คนอ่านไม่ได้ รายละเอียดดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 ความเชี่ยวชาญด้านการอ่านภาษาโบราณของนักภาษาโบราณ

ภาษา	ระดับความเชี่ยวชาญด้านการอ่าน (คน)				รวม (คน)
	มาก	ปานกลาง	น้อย	อ่านไม่ได้	
อีสาน	5	4	0	1	10
บาลี	1	6	0	3	10
ล้านนา	0	8	0	2	10
มอญ	0	1	0	9	10

4.1.2 ปัญหาการปริวรรตเอกสารโบราณ

ปัญหาในการปริวรรตเอกสารโบราณที่นักภาษาโบราณพบ แบ่งเป็น 2 ด้าน ได้แก่ ปัญหาที่มีสาเหตุจากลักษณะทางกายภาพของเอกสารโบราณ และปัญหาจากรูปอักษร ภาษา และอักษรวิธีของอักษรไทยโบราณ ดังนี้

4.1.2.1 ปัญหาที่มีสาเหตุจากลักษณะทางกายภาพของเอกสารโบราณ ได้แก่ (1) การเสื่อมสภาพของวัสดุรองรับการเขียนชำรุด แตกหัก เปราะเปื้อน ตัวอักษรลบเลือน ทำให้ยากต่อการระบุตัวอักษรและยากต่อการอ่าน โดยนักภาษาโบราณที่มีประสบการณ์ทำงานตั้งแต่ 10 ปีขึ้นไป เห็นว่าเป็นปัญหาในระดับมาก ($\bar{x} = 4.22$ SD = 0.44, 9 คน) ขณะที่นักภาษาโบราณที่มี

ประสบการณ์น้อย (2 ปี) เห็นว่าเป็นปัญหาในระดับปานกลาง และ (2) ปัญหาด้านอื่น ๆ ที่พบ อาทิ เอกสารไม่สมบูรณ์ (สูญหายบางหน้า ไม่ครบชุด หรือ เรียงสลับหน้า) ปัญหาที่เกิดจากการซ่อมแซม ผิดวิธี โดยเป็นปัญหาในระดับมาก ($\bar{x} = 4.0$ SD = 0.00, 1 คน)

4.1.2.2 ปัญหาที่มาจากรูปอักษร ภาษา และอักขรวิธีของอักษรไทยโบราณ จากการสัมภาษณ์นักภาษาโบราณสรุปประเด็นปัญหาที่พบได้ 13 ประเด็น (ตารางที่ 4.4) ส่วนใหญ่เป็นปัญหาระดับปานกลาง โดยประเด็นที่เป็นปัญหาในระดับมาก ได้แก่ อักษรโบราณไม่มีวรรณยุกต์ การเขียนคำแบบแยกตัวสะกด และการตัดคำอยู่คนละบรรทัดหรือคนละหน้า ($\bar{x} = 4.00$ SD = 0.00 และ $\bar{x} = 3.89$ SD = 0.78) ปัญหาเหล่านี้ส่งผลให้เกิดความไม่แน่ใจในการอ่านและการตีความผิดพลาด ส่วนประเด็นที่เป็นปัญหาในระดับน้อย ได้แก่ คำพ้องรูปพ้องเสียง รูปอักษรต่างจากที่เคยเห็น และการเขียนผิด ตก หรือ เกิน ทั้งนี้ เนื่องจากนักภาษาโบราณส่วนใหญ่จะมีพื้นฐานความรู้ในเรื่องที่อ่าน รู้และเข้าใจภาษาที่บันทึกในเอกสาร โบราณฉบับนั้น ๆ อยู่แล้ว รวมถึงเมื่อเกิดความไม่แน่ใจก็จะปรึกษากับผู้ที่มีความเชี่ยวชาญมากกว่า

ตารางที่ 4.4 ระดับปัญหาด้านภาษาและอักขรวิธีอักษรไทยโบราณที่ส่งผลต่อการปริวรรต

ประเด็นปัญหา	ความถี่ของระดับปัญหา (คน)					รวม (คน)	\bar{x}	SD	ระดับปัญหา
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด				
1. ลายมืออ่านยาก	1	2	7	-	-	10	3.40	0.70	ปานกลาง
2. เขียนเป็นภาษาถิ่น	-	2	6	1	-	9	3.11	0.60	ปานกลาง
3. เป็นคำศัพท์โบราณ	-	3	2	4	-	9	2.89	0.93	ปานกลาง
4. คำพ้องรูปพ้องเสียง	-	1	3	3	1	8	2.22	1.20	น้อย
5. เป็นศัพท์ที่ไม่เคยพบเห็นมาก่อน ไม่มีในพจนานุกรม	-	-	4	1	2	7	1.78	1.30	น้อยที่สุด
6. รูปอักษรที่เขียนต่างจากที่เคยเห็น	1	-	4	3	2	10	2.50	1.18	น้อย
7. เขียนอักษรโบราณอื่นปะปน	-	2	3	5	-	10	2.70	0.82	ปานกลาง
8. อักษรโบราณไม่มีวรรณยุกต์	4	1	4	-	-	9	4.00	1.00	มาก*
9. การเขียนตัวอักษรคล้ายกัน	-	2	6	2	-	10	3.00	0.67	ปานกลาง

ตารางที่ 4.4 ระดับปัญหาด้านภาษาและอักขรวิธีอักษรไทยโบราณที่ส่งผลต่อการปริวรรต (ต่อ)

ประเด็นปัญหา	ความถี่ของระดับปัญหา (คน)					รวม (คน)	\bar{X}	SD	ระดับ ปัญหา
	มากที่สุด	มาก	ปาน กลาง	น้อย	น้อย ที่สุด				
10. การเขียนผิด ตก เกิน	-	2	2	5	1	10	2.50	0.97	น้อย
11. การเขียนด้วย อักขรวิธีพิเศษ	-	3	5	1	-	9	3.22	0.67	ปานกลาง
12. การเขียนคำแยกตัว สะกด หรือ ตัดคำ แยกอยู่คนละบรรทัด หรือคนละหน้าลาน	2	4	3	-	-	9	3.89	0.78	มาก*
13. ไม่เชี่ยวชาญเนื้อหา ทุกหมวด	0	4	3	1	-	8	3.00	1.32	ปานกลาง

4.1.3 กระบวนการปริวรรตเอกสารโบราณ

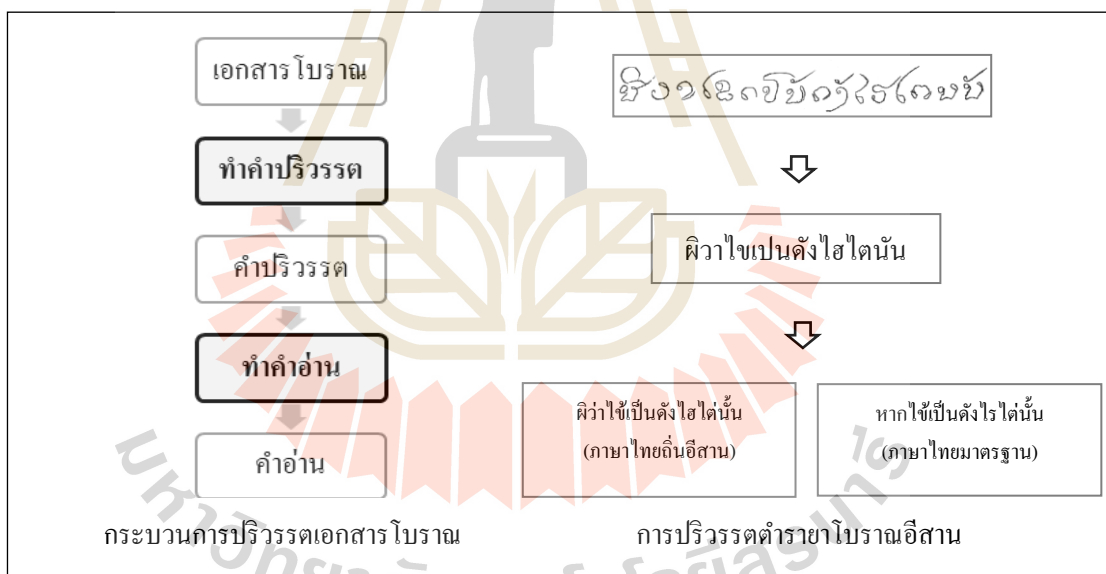
การปริวรรตเอกสาร โบราณที่นักภาษาโบราณใช้ แบ่งเป็น 2 กระบวนการ ดังนี้

4.1.3.1 การทำคำปริวรรต หมายถึง การถอดรูปอักษรโบราณเป็นอักษรไทย ปัจจุบัน ผลลัพธ์ที่ได้ในขั้นตอนนี้เรียกว่า “คำปริวรรต” โดยมีหลักเกณฑ์ คือ ถอดอักษรโบราณเป็นอักษรไทยปัจจุบันแบบตรงตัวกับอักษรโบราณในเอกสารต้นฉบับ อักษรโบราณที่เป็นอักษรแบบตัวประสม (ตัวอักษร 2 ตัว เขียนติดกัน ตัวแรกเป็นตัวอักษรเต็มตัวเขียนติดกับตัวอักษรที่ 2 ที่เป็นรูปอักษรครึ่งตัว) ถอดเป็นอักษรไทยปัจจุบัน 2 ตัว และใส่เครื่องหมายขีดบน (̄) กำกับอักษรทั้งสองเอาไว้ เช่น หิม และอักษรโบราณที่เป็นตัวควบกล้ำหรือพยัญชนะตัวตาม และตัวควบกล้ำหรือพยัญชนะตัวตามเขียนอยู่ใต้อักษรนำ เมื่อถอดเป็นอักษรไทยปัจจุบันตัวควบกล้ำหรือพยัญชนะตัวตามจะเขียนเรียงในระดับเดียวกันกับอักษรนำตามแบบอักขรวิธีประสมอักษรไทย และใส่เครื่องหมายจุด (.) ไว้ใต้อักษรนำหรือพยัญชนะต้นของตัวควบกล้ำนั้น เพื่อให้ทราบว่าพยัญชนะดังกล่าวตำแหน่งเดิมในเอกสาร โบราณต้นฉบับถูกเขียนเอาไว้ใต้อักษรนำ เช่น “กุลาว” ตำแหน่งเดิมของพยัญชนะ ล เขียนอยู่ใต้พยัญชนะ ก หรือ “โหลล” ตำแหน่งเดิมของพยัญชนะ ล เขียนอยู่ใต้พยัญชนะ ห เป็นต้น

4.1.3.2 การทำคำอ่าน หมายถึง การนำคำปริวรรตที่เขียนสะกดคำแบบ โบราณ (เขียนด้วยอักษรไทยปัจจุบันแต่ใช้ตัวสะกดตรงกับอักษรโบราณ) มาเขียนเป็นคำอ่านที่สะกดคำตามพจนานุกรม ผลลัพธ์ที่ได้เรียกว่า “คำอ่าน” การทำคำอ่านมี 2 รูปแบบคือ คำอ่านภาษาถิ่น โดยจะสะกดคำตามเสียงภาษาถิ่นที่ใช้เขียนในเอกสาร โบราณต้นฉบับ เช่น ภาษาไทยถิ่นอีสาน ภาษาเหนือ

ภาษาไทยภาคกลาง และ คำอ่านแปล ซึ่งเป็นคำอ่านที่แปลจากภาษาถิ่นเป็นภาษาไทยภาคกลาง หรือภาษาไทยมาตรฐาน และมีวัตถุประสงค์เพื่อให้คนทั่วไปอ่านเข้าใจได้ง่ายขึ้น เกณฑ์การสะกด คำอ่านใช้ตามพจนานุกรมภาษาไทย ฉบับราชบัณฑิตยสถาน ฉบับล่าสุด เป็นต้น

สำหรับกระบวนการปริวรรตคำราชาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยนั้น มักจะข้ามขั้นตอนการทำคำปริวรรตไปทำขั้นตอนการทำคำอ่าน โดยรูปแบบคำอ่านจะเป็นคำอ่าน ภาษาไทยถิ่นอีสานที่สะกดคำอ่านแบบกึ่งถอดเสียงภาษาไทยถิ่นอีสานกึ่งสะกดคำตามแบบ ภาษาไทยมาตรฐาน กล่าวคือ มีบางคำสะกดคำตามเสียงภาษาไทยถิ่นอีสาน บางคำสะกดคำตาม แบบภาษาไทยมาตรฐาน ยกตัวอย่างคำที่สะกดคำตามเสียงภาษาไทยถิ่นอีสาน เช่น “ชะเออะ” “ทกที่้น” “ไค้” เป็นต้น และคำที่สะกดคำตามแบบภาษาไทยมาตรฐาน เช่น “ปลา” “ลูกกลอน” “ซ้าง” “ร้าง” เป็นต้น ซึ่งการสะกดคำตามแบบภาษาไทยมาตรฐานจะอ่านออกเสียงไม่ตรงกับเสียง ภาษาไทยถิ่นอีสาน สรุปขั้นตอนการปริวรรตเอกสารโบราณและตัวอย่างผลการปริวรรตในแต่ละ กระบวนการ ดังแสดงในรูปที่ 4.1



รูปที่ 4.1 กระบวนการปริวรรตเอกสารโบราณ และผลลัพธ์การปริวรรตในแต่ละขั้นตอน

4.1.4 ปัญหาการปริวรรตตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย สรุปผลการศึกษาได้ 4 ประเด็น ดังนี้

4.1.4.1 ลักษณะทางกายภาพของตำรายาโบราณอีสาน

ลักษณะทั่วไปทางกายภาพของตำรายาโบราณอีสาน ได้แก่ สภาพใบลานเก่า บางลานชำรุด แตกหัก ฉีกขาด สีใบลานเปลี่ยนเป็นสีน้ำตาลเข้ม ไม่สม่ำเสมอ ตัวอักษรที่ใช้เขียนบนใบลานบางยุคกลางเดือน ไม่ชัดเจน กลมกลืนไปกับพื้นหลัง มีรอยคราบสิ่งสกปรก จุดดำ คำ และคราบมีลักษณะคล้ายคลึงกับตัวอักษร ลักษณะเหล่านี้ทำให้ยากต่อการระบุตัวอักษร ในขณะที่การชำรุดแตกหัก ใบลานขาดหายทั้งบางส่วนและขาดหายทั้งใบ ส่งผลต่อความครบถ้วนสมบูรณ์ของเนื้อหาและความถูกต้องของการปริวรรต โดยเฉพาะคำที่มีความกำกวมในการอ่านและแปลความนั้นต้องอาศัยคำบริบทมาช่วยทำให้การปริวรรตมีความถูกต้อง (รูปที่ 4.2)

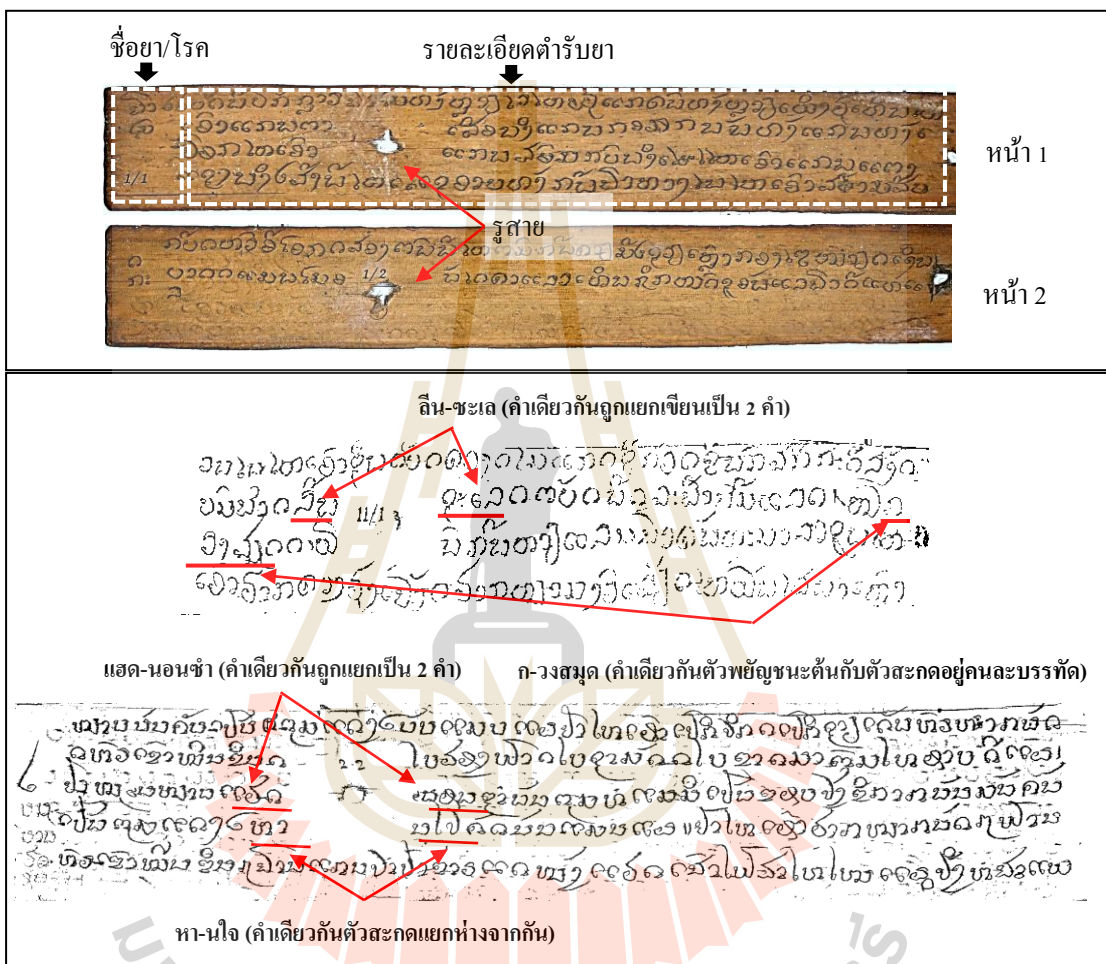


รูปที่ 4.2 ลักษณะกายภาพและการบันทึกตำรายาโบราณอีสานในเอกสารใบลาน

4.1.4.2 ลักษณะการบันทึกตำรายาโบราณอีสานในใบลานด้วยอักษรไทยน้อย

การบันทึกตำรายาโบราณอีสานจะบันทึกลงในใบลานขนาดสั้น เรียกว่า “ใบลานก้อม” (กว้าง 5-7 เซนติเมตร ยาวประมาณ 30 เซนติเมตร) พื้นที่ในการบันทึกแบ่งออกเป็น 2 ส่วน คือ พื้นที่ด้านซ้ายของใบลาน ยาวประมาณ 5 เซนติเมตร ใช้เขียนเลขหน้า และชื่อตำรับยา หรือชื่อโรค และพื้นที่ด้านขวาซึ่งเป็นพื้นที่ส่วนใหญ่ใช้เขียนรายละเอียดของตำรับยา การเขียนหรือจารตัวอักษรจะเขียนทั้ง 2 หน้าลาน หน้าละ 4 - 6 บรรทัด เขียนต่อเนื่องกันโดยไม่เว้นวรรค ดังนั้นเมื่อเขียนตัวอักษรตรงกับรู้อยาสยสนอง ตัวอักษรหรือคำที่เหลือจะถูกตัดไปเขียนข้ามรู้อยาสยสนอง และหากเขียนไปจนสุดบรรทัด หรือสิ้นสุดหน้าลานแต่ยังเขียนไม่จบคำ ผู้เขียนหรือผู้จารจะ

ตัดตัวอักษรหรือคำที่เหลือไปเขียนขึ้นบรรทัดใหม่ หรือขึ้นหน้าลานใหม่ การเขียนในลักษณะนี้ทำให้เกิดการสะกดคำแบบแยกตัวอักษร แยกคำ หรือ พยางค์ อยู่ห่างกัน ทั้งในบรรทัดเดียวกัน แยกกันอยู่คนละบรรทัด หรือ แยกกันอยู่คนละหน้า ซึ่งทำให้เกิดปัญหาในการแบ่งคำอ่าน และการตีความผิดพลาดได้ง่าย ตัวอย่างดังรูปที่ 4.3



รูปที่ 4.3 ลักษณะการบันทึกคำราชาโบราณอีสานในใบลาน

4.1.4.3 ภาษาและเนื้อหาที่บันทึกในคำราชาโบราณอีสานในใบลานด้วยอักษรไทยน้อย

ภาษาที่ใช้บันทึกคำราชาโบราณอีสานในใบลานด้วยอักษรไทยน้อย บันทึกเป็น 2 ภาษา ได้แก่ ภาษาไทยถิ่นอีสาน และภาษาบาลี ขึ้นอยู่กับวิธีการรักษาโรคหรืออาการเจ็บป่วย ซึ่งมี 3 รูปแบบ ได้แก่ รักษาด้วยสมุนไพร รักษาด้วยคาถาและพิธีกรรม (ไม่ใช่สมุนไพร) และรักษาด้วยสมุนไพรร่วมกับคาถาและพิธีกรรม ภาษาที่ใช้บันทึกจึงแตกต่างกันไปตามวิธีการรักษา และคาถามี 2 รูปแบบ ได้แก่ คาถาภาษาไทยถิ่นอีสาน และคาถาภาษาบาลีสันสกฤต ดังนั้น การจะปริวรรต

ตำราภาษาโบราณอีสานได้จึงต้องรู้จักและจำแนกได้ว่าคำใดคือคำภาษาไทยถิ่นอีสาน หรือคำใดคือภาษาบาลี โดยสิ่งที่จะช่วยให้จำแนกคือการรู้จักหรือแปลคำศัพท์นั้นได้ แต่เนื่องจากตำราภาษาโบราณอีสานมีคำศัพท์เฉพาะด้านที่คนทั่วไปไม่รู้จักหรือคุ้นเคยอยู่เป็นจำนวนมาก การปริวรรตตำราภาษาโบราณอีสานจึงมักต้องอาศัยหมอยาสมุนไพรพื้นบ้านอีสานมาช่วยปริวรรตร่วมกันกับนักภาษาโบราณ หรือเป็นที่ปรึกษาไปด้วย

4.1.3.4 การปริวรรตตำราภาษาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยในใบลานจำเป็นต้องมีทักษะหลายด้าน

จากกระบวนการปริวรรตเอกสารโบราณ 2 ขั้นตอน ได้แก่ ขั้นตอนการทำคำปริวรรต และการทำคำอ่าน ซึ่งแต่ละขั้นตอนใช้ทักษะในการดำเนินการแตกต่างกันไป คือการทำคำปริวรรตใช้ทักษะความรู้ด้านรูปอักษรและอักษรวิธีอักษรโบราณที่ใช้เขียนในเอกสารโบราณ และการทำคำอ่านอาศัยทักษะด้านความรู้ความเข้าใจในภาษาและเนื้อหาที่บันทึก อย่างไรก็ตาม การจะปริวรรตเอกสารโบราณได้อย่างมีประสิทธิภาพต้องอาศัยประสบการณ์ในการปริวรรตเอกสารโบราณอย่างต่อเนื่องเป็นระยะเวลายาวนานและเป็นจำนวนมาก ทั้งปริมาณเอกสาร อักษรภาษา และเนื้อหา สำหรับการปริวรรตตำราภาษาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยจะใช้ทักษะสำคัญ 5 ด้าน (รูปที่ 4.4) ได้แก่ (1) รู้จักอักษรไทยน้อยและอักษรธรรมอีสาน (2) รู้และเข้าใจภาษาไทยถิ่นอีสานและบาลี (3) มีความรู้เกี่ยวกับตำราภาษาโบราณอีสาน (4) มีประสบการณ์ในการปริวรรตอักษรไทยน้อยและอักษรธรรมอีสานมาเป็นเวลานาน และ (5) มีประสบการณ์ในการปริวรรตเอกสารโบราณหลายหมวดหมู่เนื้อหา



รูปที่ 4.4 แผนภาพทักษะสำหรับปริวรรตตำราภาษาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย

4.2 การศึกษาโครงสร้างการเขียนตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทย น้อยที่ส่งผลต่อประสิทธิภาพในการปริวรรต

การวิเคราะห์โครงสร้างการเขียนตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อยในการศึกษาครั้งนี้ ทำการวิเคราะห์จากหนังสือตำรายาที่ปริวรรตมาจากตำรายาโบราณอีสานที่บันทึกไว้ในใบลาน จำนวน 2 เล่ม ได้แก่ ตำรายาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 5 และ ตำรายาวัดท่าม่วง อำเภอสหัสขันธ์ จังหวัดร้อยเอ็ดเล่ม 2 โดยทั้ง 2 เล่มนี้ปริวรรตโดยงานอนุรักษ์เอกสารโบราณ สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม คัดเลือกมาศึกษา เฉพาะฉบับที่บันทึกด้วยอักษรไทยน้อยจำนวนทั้งสิ้น 11 ผูก 459 หน้าลาน รายละเอียดดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 รายละเอียดตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อยที่นำมาศึกษา

ชื่อตำรายา	วัดมหาชัย เล่ม 5							วัดท่าม่วง เล่ม 2				รวม
	1	2	3	4	5	6	8	8	9	10	11	
ฉบับที่ (ผูก)	1	2	3	4	5	6	8	8	9	10	11	11
จำนวนลาน	5	36	14	16	11	51	37	15	12	20	16	223
จำนวนหน้า	10	71	28	31	21	102	73	29	23	39	32	459
จำนวนบรรทัด/หน้า	4	4	4	5	4	4, 5	4	4	4	4	6	

ผลการศึกษาแบ่งได้เป็น 2 ด้าน ได้แก่ องค์ประกอบเนื้อหาที่บันทึกในตำรับยาโบราณอีสาน และโครงสร้างการเขียนตำรายาโบราณอีสาน รายละเอียดมีดังนี้

4.2.1 องค์ประกอบด้านเนื้อหาที่บันทึกในตำรับยาอีสานโบราณ

องค์ประกอบด้านเนื้อหาที่เขียนในตำรับยาอีสานโบราณ แบ่งเป็น 12 กลุ่มหลัก ดังนี้

4.2.1.1 ชื่อยา/โรค/อาการ ได้แก่ ชื่อตำรับยา หรือ ชื่อโรค หรือ อาการของโรค เช่น สันนิบาตไฟ หมากไม้แลง หมากขบ (หมากกัด) หนาวโน สะเออะ (สะอึก) หนหวย (รำคาญ/หงุดหงิด) เป็นต้น

4.2.1.2 ชื่อสมุนไพร ได้แก่ ชื่อสมุนไพรที่ใช้ในการรักษาโรค/อาการ แบ่งเป็น 4 ประเภท ได้แก่

1) พืชวัตถุ หมายถึง ชื่อพืชและส่วนประกอบของพืชที่ใช้ทำยา เช่น ผักหวาน เข้าหมื่น (ขมิ้น) เล็บแมว (เล็บเหยี่ยว) ดอกเข็ม หมากเคื่อ (มะเคื่อ) ฮากไผ่บ้าน จี่ชี (ชัน) เป็นต้น

2) ธาตุวัตถุ หมายถึง ชื่อแร่ธาตุที่ใช้ทำยา เช่น ชิน (ชิน) หินแห่ (กรวด) มาด (กำมะถัน) ขาง (เหล็กชนิดหนึ่ง) เป็นต้น

3) สัตว์วัตถุ หมายถึง ชื่อสัตว์ หรือส่วนประกอบของสัตว์ที่ใช้ทำยา เช่น คันคาก (คางครก) หอยจูป (หอยขม) บิงูเหลือม (ดิ้งูเหลือม) หนั่งแซ่ (หนั่งจระเข้) คราบแมงมุม กระดูกของสัตว์ต่าง ๆ เป็นต้น

4) กระจายยา หมายถึง เครื่องแทรกยา น้ำ หรือของเหลวที่ใช้ละลายยา เช่น น้ำผึ้ง (น้ำผึ้ง) น้ำฝน น้ำอุ่น เป็นต้น

4.2.1.3 สัดส่วน/ปริมาณที่ใช้ หมายถึง การระบุสัดส่วน ปริมาณ หรือ น้ำหนัก สมุนไพรแต่ละชนิดที่นำมาใช้ปรุงยาแต่ละตำรับ เช่น ก้อนนึ่ง (1 ก้อน) คู่หนึ่ง (1 คู่) ซีก อันแลอัน (อย่างละอัน) 3 ฟาน บาท 1 (1 บาท) เป็นต้น

4.2.1.4 วิธีไปเก็บสมุนไพร หมายถึง พิธีกรรม วิธีการ หรือ ข้อกำหนดเมื่อจะเก็บ สมุนไพรมาใช้ทำยา เช่น ให้ไปเก็บสมุนไพรทางทิศตะวันตก ตอนเก็บสมุนไพรให้กลั่นหายใจ วัน และเวลาที่ให้ไปเก็บสมุนไพร เป็นต้น

4.2.1.5 วิธีเตรียมยา หมายถึง วิธีการนำยามาแปรรูปหรือปรุงด้วยวิธีการต่าง ๆ เช่น ฝน แซ่ (แซ่) หลาม (ใส่กระบอก) สม (รม) บด เผา ตากแดด เป็นต้น โดยรูปแบบผลิตภัณฑ์ที่ได้ เช่น น้ำ ผง (ฝุ่น) ลูกกลอน เป็นต้น

4.2.1.6 วิธีใช้ยา หมายถึง วิธีการนำยามาใช้รักษาอาการต่าง ๆ เช่น แห่ ทา อาบ สม (รม) สูบ กิน เป้า แจะ (แตะ) จ้ำ แต้้ม แทะ ดบหัว ตั่งหัว พุด (พน) เป็นต้น

4.2.1.7 ผู้ป่วย หมายถึง ชื่อเรียกผู้ป่วยที่เป็น มนุษย์ ได้แก่ เด็กน้อย แม่หญิง (ผู้หญิง) พ่อชาย (ผู้ชาย) หรือ สัตว์ที่ป่วย เช่น ช้าง (ช้าง) ม้า วัว ควาย เป็นต้น โดยอาจมีการบอกชื่ออวัยวะที่มีอาการเจ็บป่วยด้วย เช่น หัว หู แขน ขา ท้อง เป็นต้น

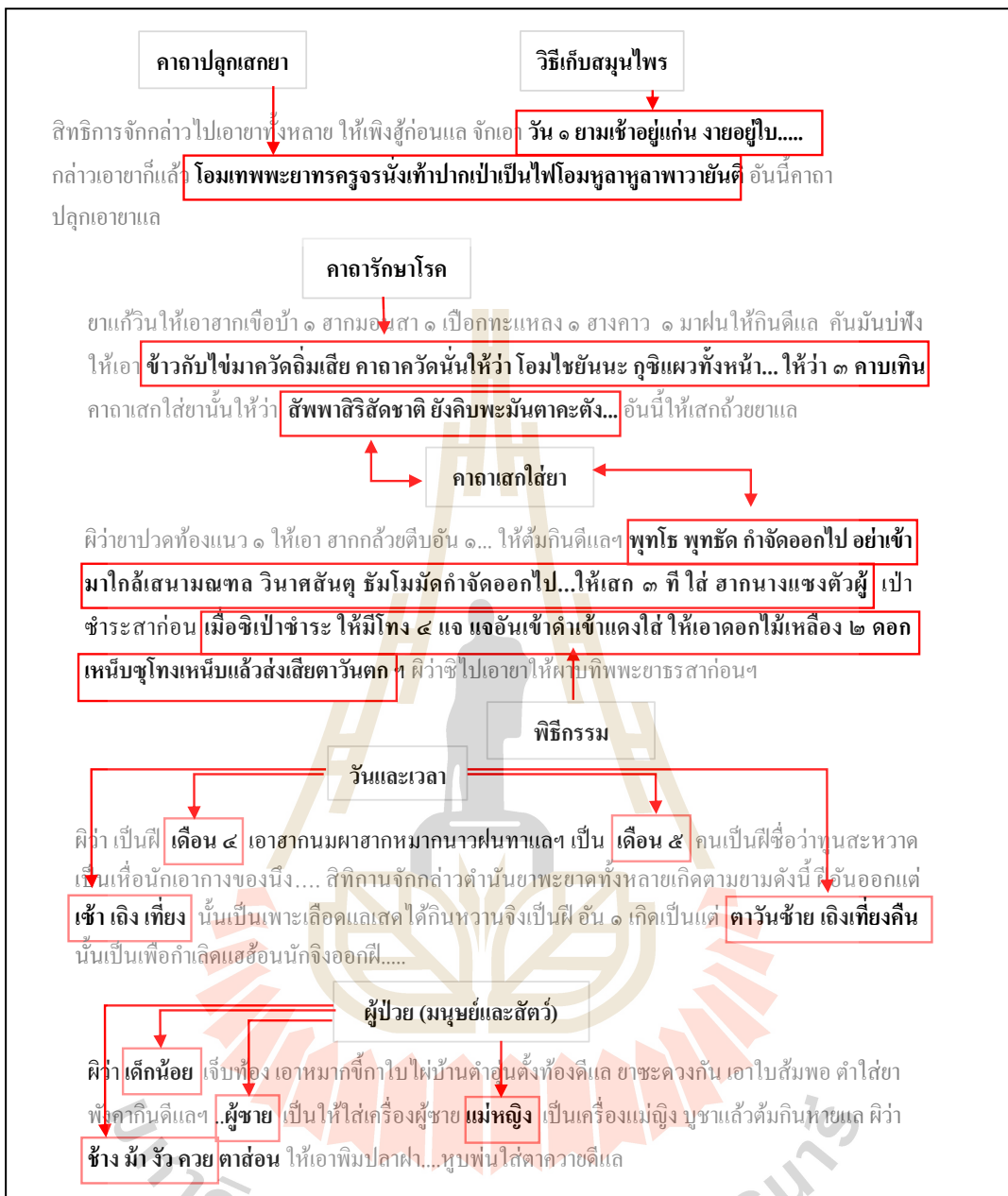
4.2.1.8 วันและเวลา หมายถึง ชื่อเรียก วัน เดือน หรือ เวลา ดังนี้

1) ชื่อเรียกวัน เช่น วัน๑ (จันทร์) วัน๒ (อังคาร) วัน๓ (พุธ) วันปะหัส (พฤหัสบดี) วันทิตย (อาทิตย์) วันเดือนดับ (ข้างแรม) วันเพ็ญ (วันเพ็ญ/ข้างขึ้น) เป็นต้น

2) ชื่อเรียกเดือน เช่น เดือนเจียง (เดือนอ้าย) เดือนยี่ (เดือนสอง) เดือนสาม เป็นต้น

3) ชื่อเรียกช่วงเวลา เช่น เช้า (เช้า) เที่ยง เมื่อแลง (ตอนเย็น) ก่อนงาย (ก่อนเที่ยง) ดาวนชาย (ตะวันตก) เป็นต้น

4.2.1.9 คาถา หมายถึง คำกล่าวแบบร้อยกรองที่ใช้เมื่อเก็บสมุนไพร ปลูกเสกยา เตรียมยา/ปรุงยา ใช้ยา หรือใช้รักษาโรค โดยคาถามี 2 รูปแบบคือ คาถาภาษาไทยถิ่นอีสาน และคาถาภาษาบาลี



รูปที่ 4.6 ตัวอย่างองค์ประกอบด้านเนื้อหาในตำรายาโบราณอีสาน

4.2.2 โครงสร้างการเขียนตำรายาโบราณอีสาน

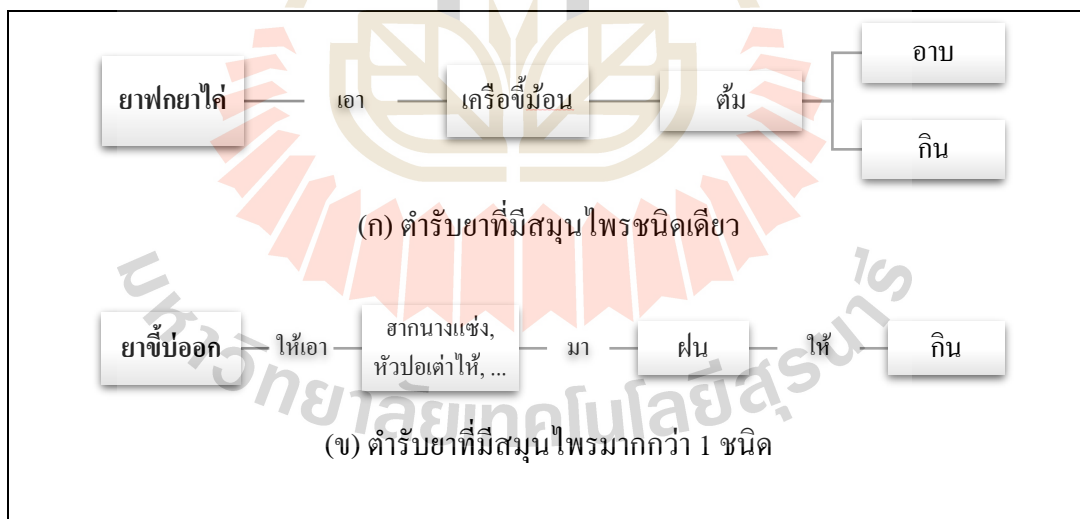
การเขียนตำรายาโบราณอีสานมักเริ่มต้นด้วยการกล่าวถึงสิ่งศักดิ์สิทธิ์ เช่น ลิตทิกการริยะ... ลิตทิกกาล... แล้วจึงตามด้วยเนื้อหาตำรับยาแต่ละตำรับ โดยโครงสร้างการเขียนบรรยายเนื้อหาในตำรายาโบราณอีสาน แบ่งเป็น 2 ประเภท คือ โครงสร้างเนื้อหาประเภทตำรับยา และเนื้อหาอื่น ๆ ที่ไม่ใช่ตำรับยา เนื้อหาประเภทตำรับยา แบ่งเป็น 3 รูปแบบ ได้แก่ ตำรับยาประเภทใช้

สมุนไพรในการรักษา ตำรับยาประเภทใช้สมุนไพรร่วมกับพิธีกรรมและคาถาในการรักษา และ ตำรับยาประเภทที่ใช้เฉพาะคาถาในการรักษา ลักษณะ โครงสร้างของเนื้อหาแต่ละประเภทมีดังนี้

4.2.2.1 โครงสร้างการเขียนบรรยายเนื้อหาประเภทตำรับยา แบ่งเป็น 3 ประเภท ดังนี้

1) ตำรับยาประเภทใช้เฉพาะสมุนไพรในการรักษา มีลักษณะ โครงสร้าง การเขียนคล้ายคลึงกัน คือ บอกชื่อยา ชื่อโรค หรือ อาการเจ็บป่วย หรืออาการของโรค อวัยวะที่เป็น โรค บางตำรับอาจบอกเฉพาะชื่อยา หรือชื่อโรค บางตำรับจะบอกเฉพาะอาการ แล้วจึงบอกรายการสมุนไพรที่ใช้รักษาพร้อมสัดส่วนปริมาณที่ใช้ วิธีเตรียมยา วิธีใช้ยา โดยวิธีใช้ยาบางตำรับอาจมี วิธีใช้ต่างกันขึ้นอยู่กับเพศของผู้ป่วย ตำรับยาสำหรับรักษาโรคหนึ่งโรค อาจมีตำรับย่อย ๆ มากกว่า 1 ขนาน โดยมีขนานหลัก และขนานทางเลือก หรือ บางตำรับอาจมีขนานหลัก และขนานทางเลือก สำหรับใช้เมื่อยาขนานหลักใช้ไม่ได้ผล หรือใช้แล้วอาการไม่ดีขึ้น หรือมีอาการบางประการเกิดขึ้น หลังจากใช้ยาขนานหลักไปแล้ว ตำรับยาบางตำรับรักษาโรคเพียงโรคเดียว บางตำรับรักษาได้หลายโรค ตำรับยาบางตำรับใช้รักษาสัตว์ โดยสรุปลักษณะ โครงสร้างตำรับยาประเภทใช้สมุนไพรในการรักษา มีดังนี้

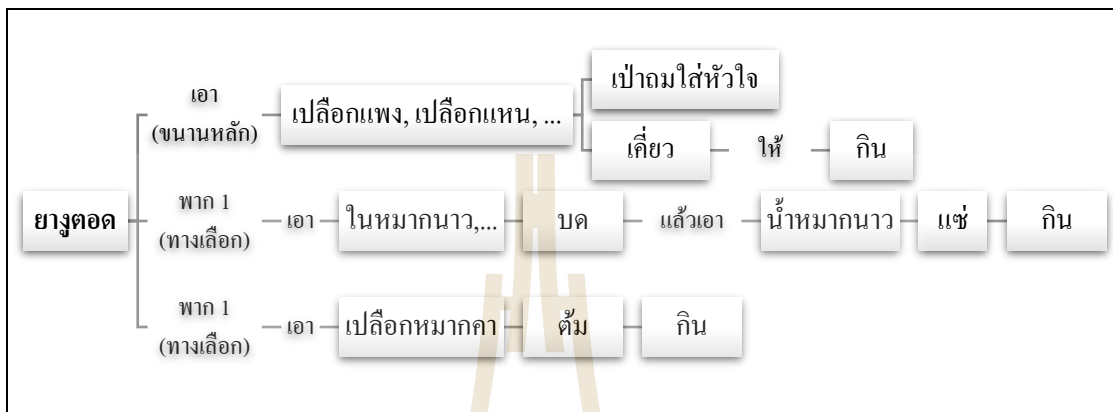
1.1) รักษาได้ 1 โรค มี 1 ขนาน ตัวอย่างผังรูปที่ 4.7



รูปที่ 4.7 ตัวอย่างโครงสร้างตำรับยาประเภทใช้สมุนไพรที่รักษาได้ 1 โรค มี 1 ขนาน

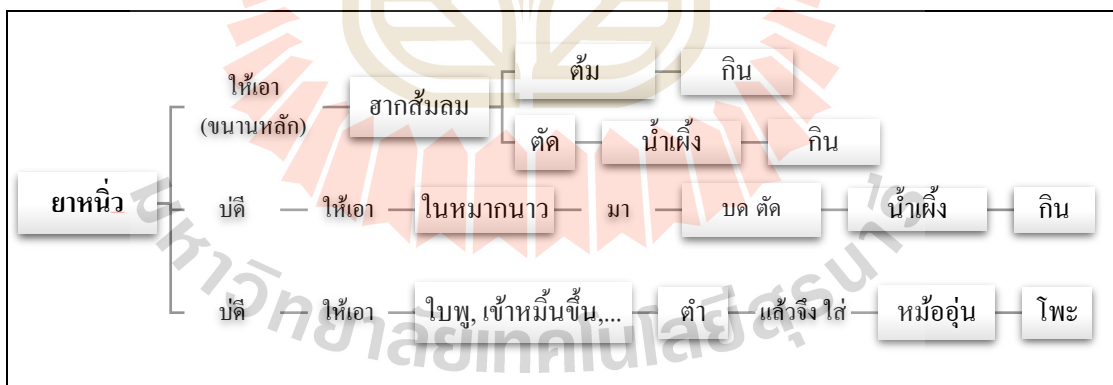
1.2) รักษาได้ 1 โรค มีขนานหลัก และขนานทางเลือก ตัวอย่าง

ดังรูปที่ 4.8



รูปที่ 4.8 ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค มีขนานหลัก และขนานทางเลือก

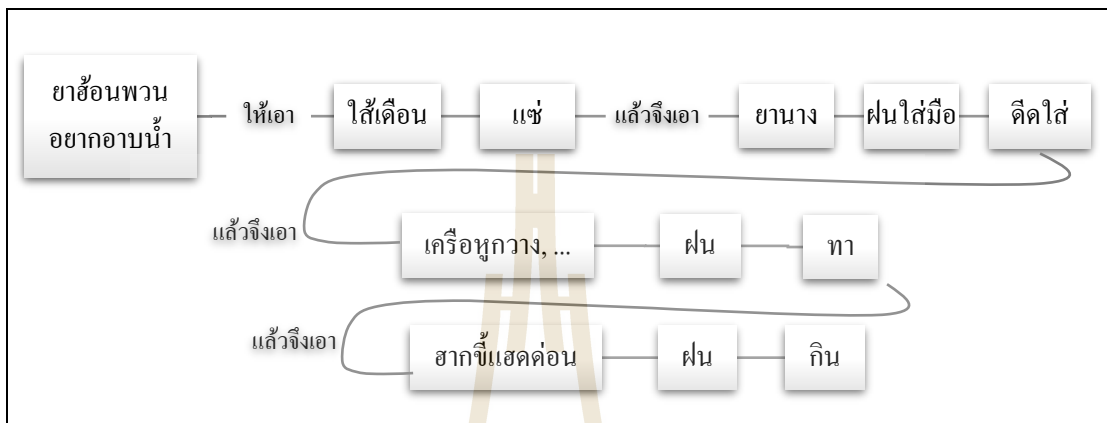
1.3) รักษาได้ 1 โรค มีขนานหลัก และขนานทางเลือก ขนานทางเลือกใช้เมื่อขนานหลักใช้ไม่ได้ผล ตัวอย่างดังรูปที่ 4.9



รูปที่ 4.9 ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค มีขนานหลัก และขนานทางเลือก สำหรับใช้เมื่อขนานหลักใช้ไม่ได้ผล)

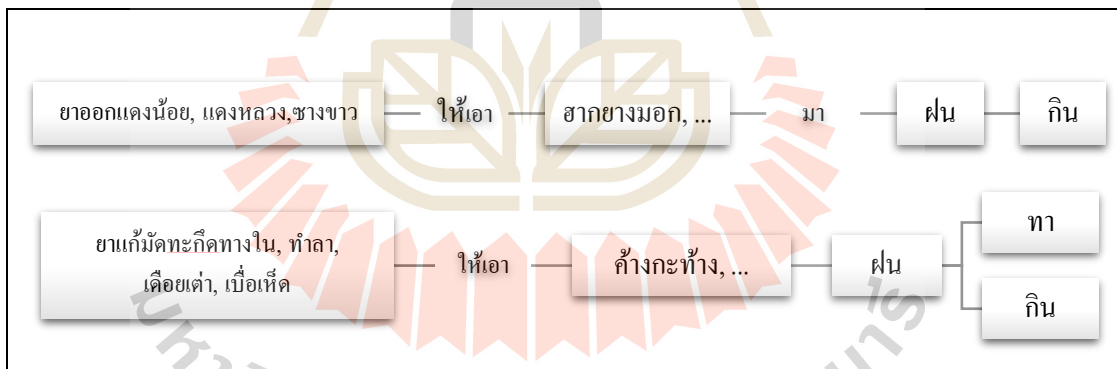
1.4) รักษาโรค 1 โรค และมีขนานย่อย ๆ ให้ใช้ตามลำดับ ตัวอย่าง

ดังรูปที่ 4.10



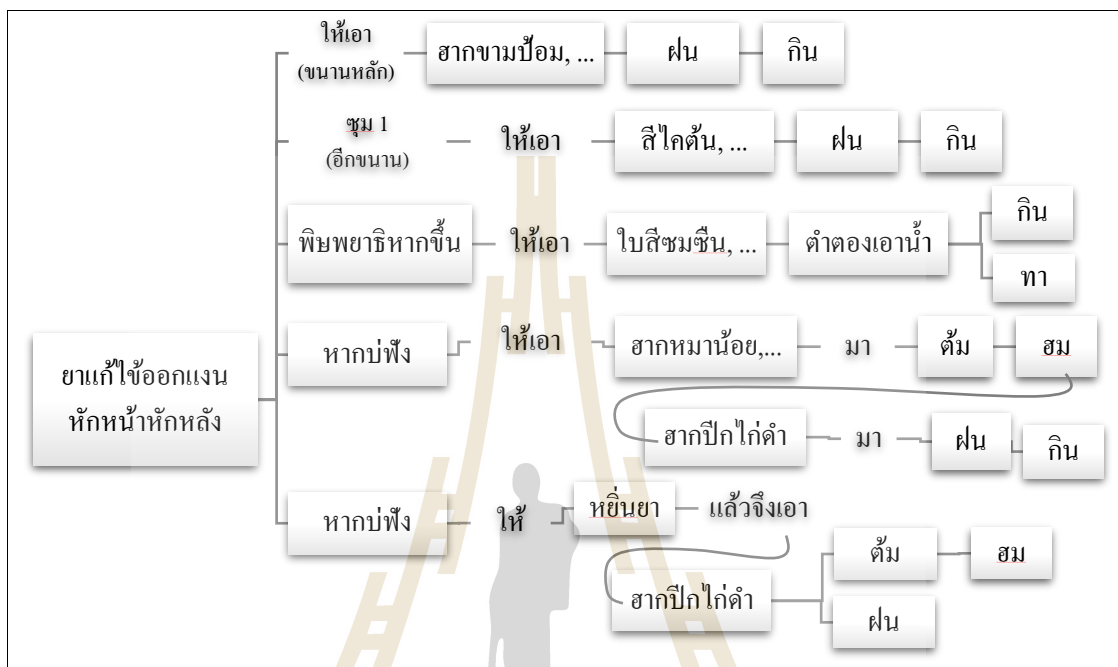
รูปที่ 4.10 ตัวอย่าง โครงสร้างตำรับยาสมุนไพรที่รักษาโรค 1 โรค และขนานย่อยใช้ตามลำดับ

1.5) รักษาโรคหลายโรคด้วยขนานเดียวกัน ตัวอย่างดังรูปที่ 4.11



รูปที่ 4.11 ตัวอย่าง โครงสร้างตำรับยาสมุนไพรที่รักษาโรคได้หลายโรคด้วยขนานเดียวกัน

1.6) รักษาได้ 1 โรค มีหลายขนาน แต่ละขนานใช้ตามอาการหรือใช้เมื่อขนานก่อนหน้ารักษาไม่ได้ผล ตัวอย่างดังรูปที่ 4.12

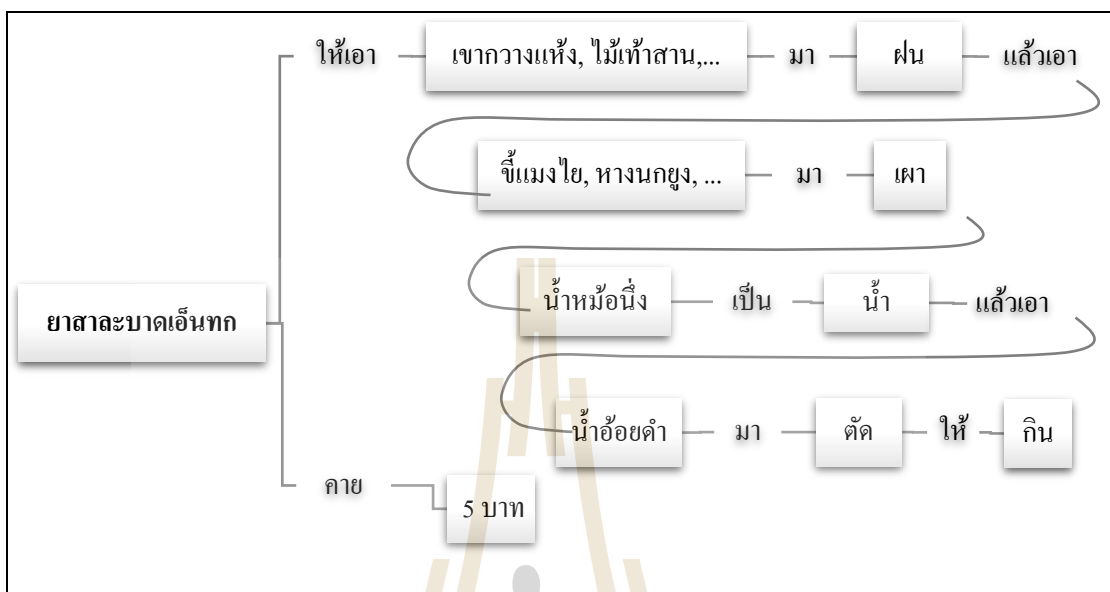


รูปที่ 4.12 ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่มีหลายขนาน แต่ละขนานใช้ตามอาการหรือใช้เมื่อขนานก่อนหน้าใช้ไม่ได้ผล

1.7) ตำรับยาที่มีค่ารักษา (ค่าย) ตัวอย่างดังรูปที่ 4.13

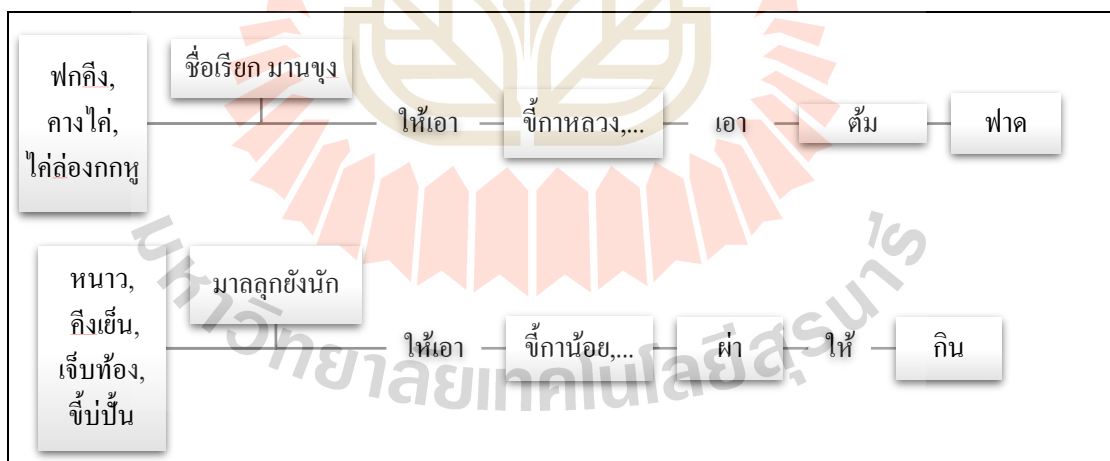


รูปที่ 4.13 ตัวอย่าง โครงสร้างตำรับยาสมุนไพรที่มีค่ารักษา (ค่าย)



รูปที่ 4.13 ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่มีค่ารักษา (คาย) (ต่อ)

1.8) ตำรับยาที่บอกอาการของโรคและชื่อโรค ตัวอย่างดังรูปที่ 4.14



รูปที่ 4.14 ตัวอย่างโครงสร้างตำรับยาสมุนไพรที่บอกอาการของโรคและชื่อโรค

1.9) คำรับยาที่บอกชื่อโรค อาการ และชื่อเรียกอาการ ตัวอย่าง

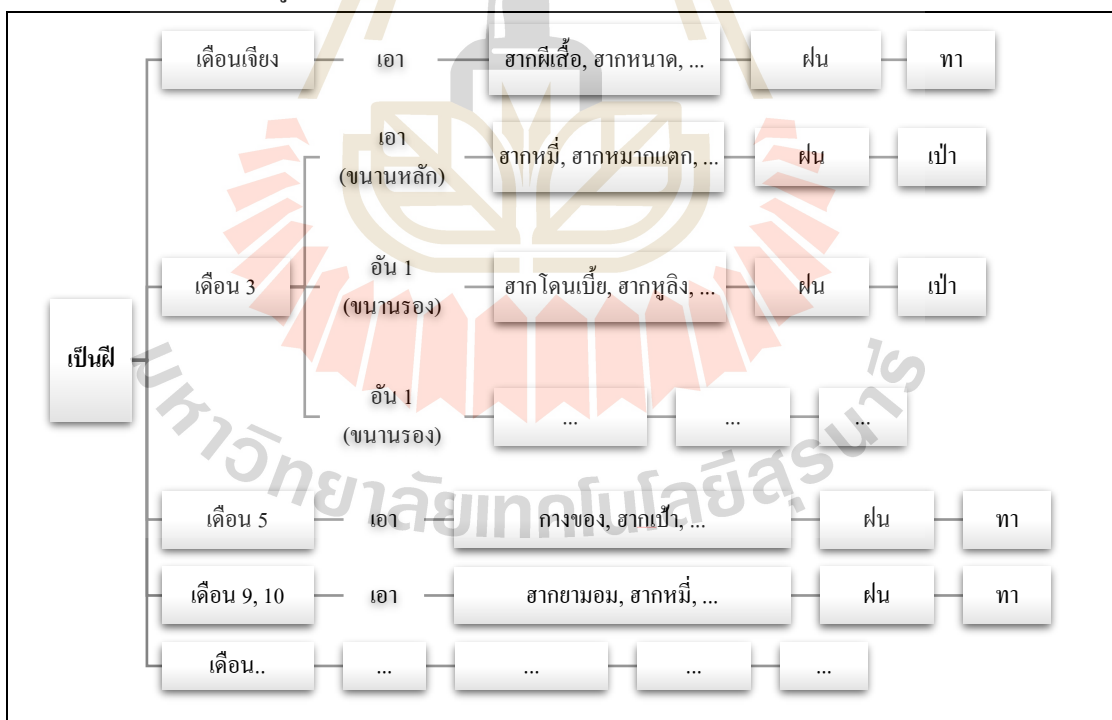
ผังรูปที่ 4.15



รูปที่ 4.15 ตัวอย่างโครงสร้างคำรับยาสมุนไพรที่บอกชื่อโรค อาการ และชื่อเรียกอาการ

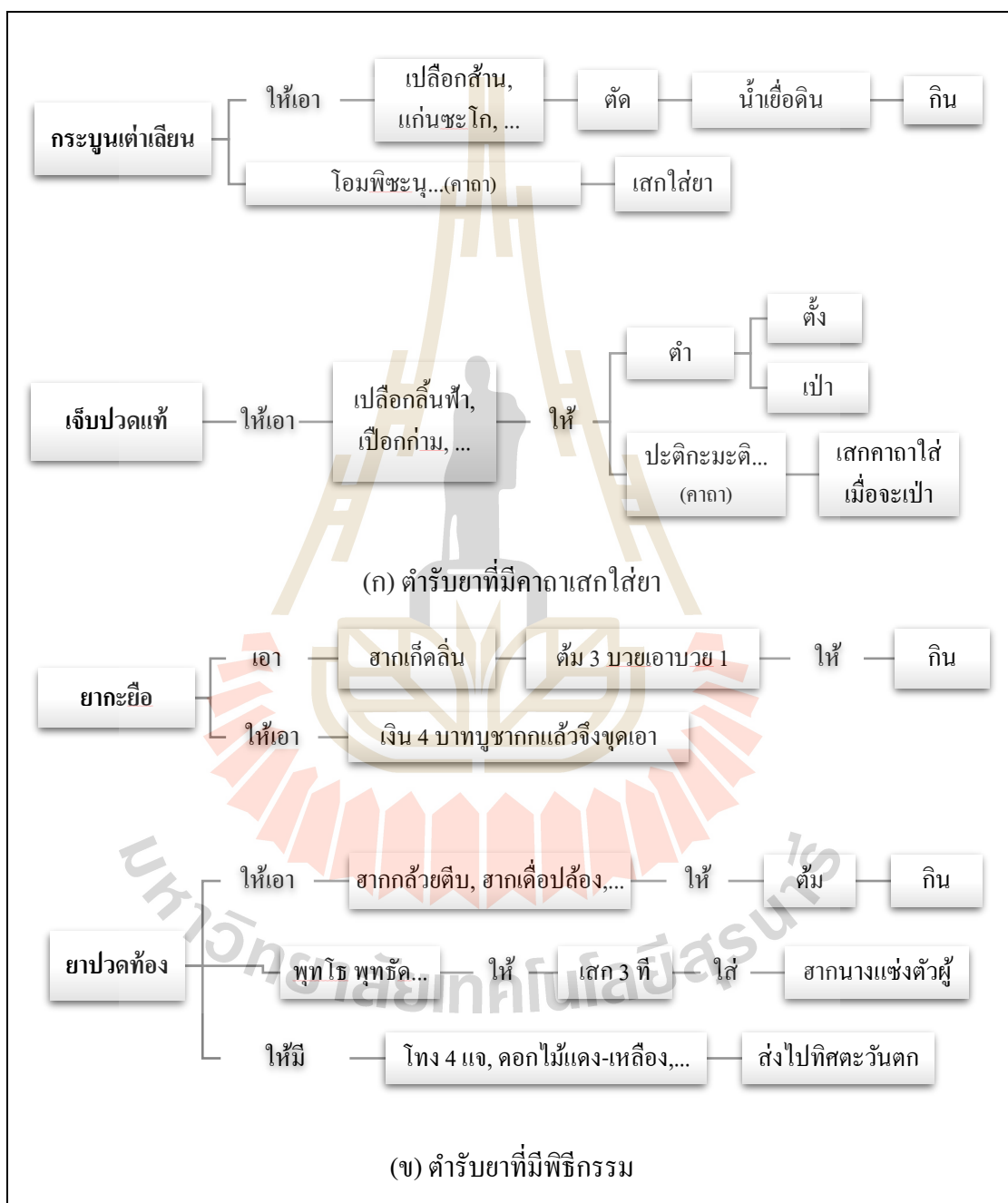
1.10) คำรับยาที่รักษา 1 โรค มีหลายขนาน แต่ละขนานใช้ตาม

เดือนที่ป่วย ตัวอย่างผังรูปที่ 4.16



รูปที่ 4.16 ตัวอย่างโครงสร้างคำรับยาสมุนไพรที่รักษา 1 โรค มีหลายขนานให้ใช้ตามเดือนที่ป่วย

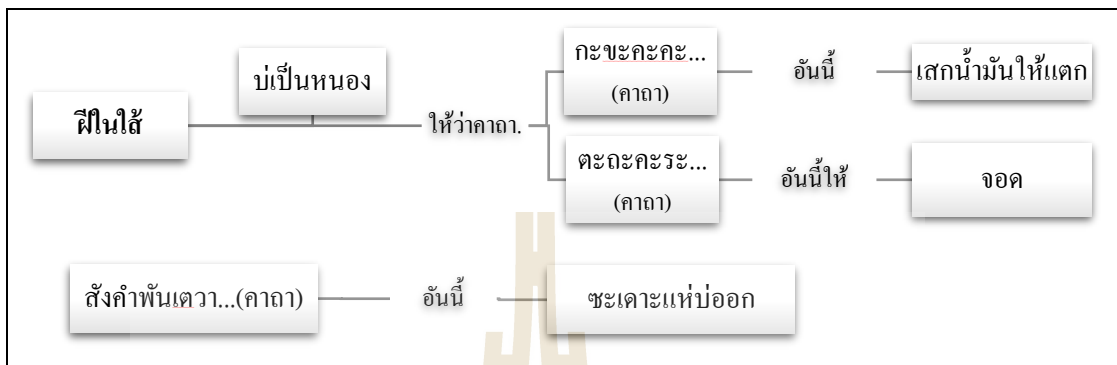
2) คำรับยาประเภทใช้สมุนไพรร่วมกับพิธีกรรม หรือคาถาในการรักษาโรค เช่น เสกคาถาใส่สมุนไพรเมื่อตอนไปเก็บ ขณะเตรียมยา ประโยก ก่อน หรือเมื่อจะใช้ยา หรือมีพิธีกรรมร่วมกับการรักษา เช่น การเซ่นไหว้โดยการทำกระทงใส่เครื่องบูชา ตัวอย่างดังรูปที่ 4.17



รูปที่ 4.17 ตัวอย่างโครงสร้างคำรับยาที่ใช้สมุนไพรร่วมกับพิธีกรรมหรือคาถาในการรักษา

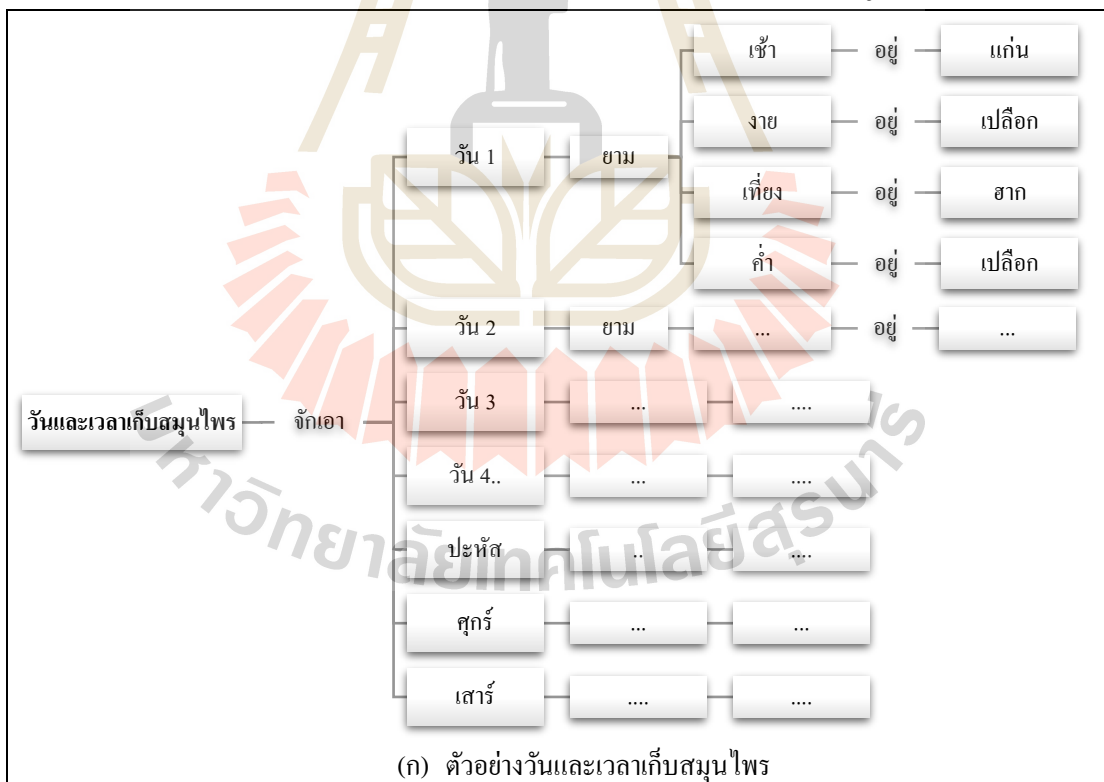
3) คำรับยาประเภทรักษาโรคด้วยคาถา (ไม่ใช่สมุนไพร) ตัวอย่างดังรูปที่

4.18



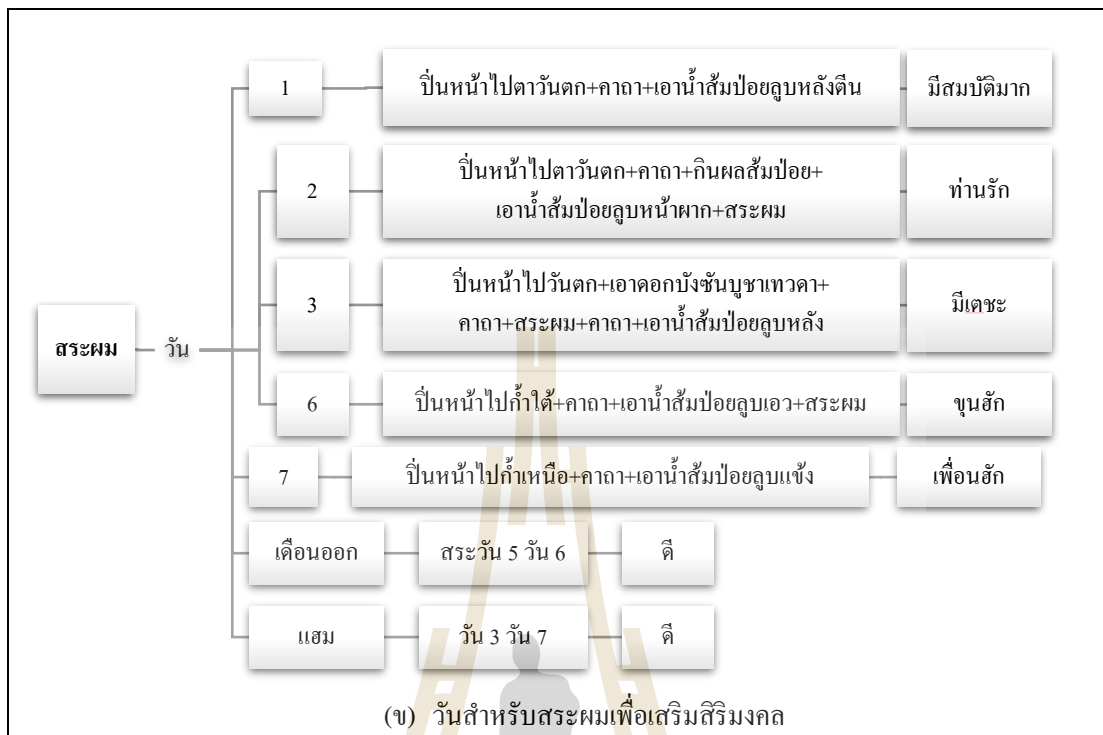
รูปที่ 4.18 ตัวอย่างโครงสร้างคำรับยาประเภทรักษาโรคด้วยคาถา

4.2.2.2 เนื้อหาประเภทอื่น ๆ ได้แก่ เนื้อหาที่ไม่ใช่คำรับยาสำหรับรักษาโรค เช่น วันเวลาในการไปเก็บสมุนไพร พิธีกรรมเสริมความเป็นสิริมงคล ตัวอย่างดังรูปที่ 4.19



(ก) ตัวอย่างวันและเวลาเก็บสมุนไพร

รูปที่ 4.19 ตัวอย่างโครงสร้างเนื้อหาประเภทอื่น ๆ ในตำรายาโบราณอีสาน



รูปที่ 4.19 ตัวอย่างโครงสร้างเนื้อหาประเภทอื่น ๆ ในตำรายาโบราณอีสาน (ต่อ)

4.2.3 การจัดกลุ่มคำศัพท์ที่ใช้เขียนในตำรายาโบราณอีสาน

ผลจากการจำแนกโครงสร้างเนื้อหาในตำรายาโบราณอีสาน สามารถนำมาจัดกลุ่มคำศัพท์ภาษาถิ่นอีสานที่ใช้เขียนแบ่งเป็น 2 หมวด ได้แก่ หมวดคำรับยา และหมวดอื่น ๆ ดังแสดงในรูปที่ 4.20 ตัวอย่างคำศัพท์ในแต่ละหมวดมีดังนี้

4.2.3.1 คำรับยา แบ่งเป็น 10 หมวดย่อย ได้แก่

- 1) ยา แบ่งเป็น 2 หมวดย่อย ได้แก่
 - 1.1) ยารักษามนุษย์ เช่น ยาฟก ยามะเร็ง (ยามะเร็ง) ยาฝิโนไล้ ยาแก้พวนใน (ยาแก้มวนท้อง) ยาชักพิษ เป็นต้น
 - 1.2) ยารักษาสัตว์ เช่น ยาหนอนกินจัวกินควย (หนอนกินวัวควาย) ยาจัวควยตาส่อน (ยาวัวควยตาส่อน) เป็นต้น
- 2) อาการป่วย/ชื่อโรค คำศัพท์ในหมวดนี้ เช่น ไข้ ฮ้อน (ร้อน) ปวดหัว เจ็บท้อง (ปวดท้อง) ล่องแก้วไฟ สารบาด คางเข้ากับ (คางแข็ง) กะยือ (หืด) เป็นต้น
- 3) สมุนไพร แบ่งเป็น 4 หมวดย่อย ได้แก่
 - 3.1) พืชวัตถุ แบ่งเป็น 3 หมวดย่อย ได้แก่

3.1.1) ชื่อพืชวัตถุ เช่น เก็ดลั่นหลวง (เก็ดลั่นหลวง) เลี้ยวเงิน
เลียง เป็นต้น

3.1.2) ชื่อส่วนประกอบของพืชวัตถุ เช่น ใบ ต้น เปลือก (เปลือก)
ฮาก (ราก) จีชี (ชัน) ตัวอย่างเช่น ฮากฝักอืดู (รากคั้นแมงลัก) ฮากดอกพุด (รากคั้นพุด) หมากพิกกัน
ชัน (ผลพริกชี้ฟ้า) เป็นต้น

3.1.3) ลักษณะนามพืชวัตถุ เช่น กำ ฝาน หัว ดอก ใบ เป็นต้น

3.2) สัตว์วัตถุ แบ่งเป็น 3 หมวดย่อย ได้แก่

3.2.1) ชื่อสัตว์วัตถุ เช่น งู ควย (ควาย) ปลา (ปลา) ไก่ดำ คันคา
(กลางครก) กะท่าง (กึ่งกำ) เป็นต้น

3.2.2) ชื่อส่วนประกอบสัตว์วัตถุ เช่น ลูก (กระดูก) เขา งา ขน
ไก่ บี (น้ำดี) หนั่ง จี้ควยด่อน (จี้ควายเผือก) บิงูเหลื่อม (ดิงูเหลื่อม) เป็นต้น

3.2.3) ลักษณะนามสัตว์วัตถุ เช่น โด (ตัว) คาบ (คราบ) ฮัง (รัง)
เป็นต้น

3.3) ธาตุวัตถุ แบ่งเป็น 2 หมวดย่อย ได้แก่

3.3.1) ชื่อธาตุวัตถุ เช่น หินแห่ (หินกรวด) มาด (กำมะถัน) ขาง
(เหล็กชนิดหนึ่ง) คำ (ทองคำ) เงิน เป็นต้น

3.3.1) ลักษณะนามของธาตุวัตถุ เช่น ก้อน เป็นต้น

3.4) กระจายยา เช่น น้ำผึ้ง (น้ำผึ้ง) น้ำ น้ำมวก (น้ำแซ่ข้าว) น้ำฝน
เหล่า เป็นต้น

4) วิธีเตรียมยา/ปรุงยา แบ่งเป็น 5 หมวดย่อย ได้แก่

4.1) วิธีเตรียมยา/ปรุงยา เช่น ฝน ค่อม ตำ ดาก หลาม (เผาใส่กระบอ)
เป็นต้น

4.2) จำนวน/ปริมาณ ได้แก่ จำนวนแบบตัวเลข เช่น 1 2 5 7 จำนวน
แบบข้อความ หนึ่ง ห้า ห้าร้อย (500) 2ร้อย (200) และจำนวนแบบคำบอกปริมาณ เช่น หน่อยหนึ่ง
สน้อย (สักหน่อย) ท่อ ประมาณ หยิบ เป็นต้น

4.3) หน่วยวัดและชั่งตวง เช่น บาท ตำลึง ช้อมือ เป็นต้น

4.4) อุปกรณ์ปรุงยา เช่น ถ้วย แพฆาว (ผ้าขาว) หม้อ บั้ง (กระบอ)
เป็นต้น

4.5) รูปแบบผลิตภัณฑ์ เช่น ฝุ่น (ผง) น้ำ ลูกกลอน เป็นต้น

5) วิธีใช้ยา แบ่งเป็น 2 หมวดย่อย ได้แก่

- 5.1) วิธีใช้ยา เช่น กิน สม (รม) อาบ ทา โปะะ (โป๊ะ) ผิว (ลูบ) เป็นต้น
- 5.2) คำข้อห้ามในการใช้ยา ได้แก่ ละล้า เป็นต้น
- 6) สรรพคุณยา เช่น กินดี ทาดี อาบดี เป็นต้น
- 7) คำรักษา เช่น คาย คำ (ทองคำ) เงิน ผ้ำ
- 8) คาถา แบ่งเป็น 3 หมวดย่อย ได้แก่
- 8.1) คาถารักษาโรค เช่น สังกำพันเตวา...
- 8.2) คาถาเสกใส่สมุนไพโร เช่น กะชะกะ...
- 8.3) คาถาเก็บสมุนไพโร เช่น พุทโธ พุคทัด...
- 9) ผู้ป่วย แบ่งเป็น 2 หมวดย่อย ได้แก่
- 9.1) มนุษย์ แบ่งเป็น 3 หมวดย่อย ได้แก่
- 9.1.1) ชื่อ/สรรพนามมนุษย์ เช่น เขา เจ้า ผู้ใช้ เฮา เด็กน้อย เป็นต้น
- 9.1.2) คำเรียกเพศมนุษย์ เช่น พ่อชาย (ผู้ชาย) แม่หญิง (ผู้หญิง)
- 9.1.3) ชื่ออวัยวะมนุษย์ เช่น แข้ง ขา แขน มือ หัว ท้อง เป็นต้น
- 9.2) สัตว์ แบ่งเป็น 3 หมวดย่อย ได้แก่
- 9.2.1) ชื่อสัตว์/สรรพนามสัตว์ เช่น ช้าง (ช้าง) ม้า จั้ว (วัว) ควย (ควาย) เป็นต้น
- 9.2.2) คำเรียกเพศสัตว์ เช่น ผู้ (ตัวผู้) แม่ (เมีย) เป็นต้น
- 9.2.3) ชื่ออวัยวะสัตว์ เช่น เขา งา หนั่ง ก้าง ลูก (กระดูก) เป็นต้น
- 10) พิธีกรรมรักษาโรค แบ่งเป็น 2 หมวดย่อย ได้แก่
- 10.1) เครื่องประกอบพิธีกรรม เช่น โทง (กระทง) ส่วย (กรวย) ตา แล่วหน้าจั่ว (ธง) เข้าดำ (ข้าวดำ) เข้าแดง (ข้าวแดง) เป็นต้น
- 10.2) พิธีกรรม เช่น บันสูปคน (บันรูปคน) เสก เป็นต้น
- 4.2.3.2) อื่น ๆ แบ่งเป็น 9 หมวดย่อย ได้แก่
- 1) วันเวลา แบ่งเป็น 3 หมวดย่อย ได้แก่
- 1.1) คำเรียกเวลา เช่น มือแลง (ยามค่ำ) มือเช้า (ยามเช้า) งาม (สาย) เทียง คำตะวันชาย (ตะวันขึ้น) เป็นต้น
- 1.2) คำเรียกวัน เช่น วันหนึ่ง (วันจันทร์) วันอังคาร วันปะหัดส (พฤหัสบดี) วันศุกร์ วันเสาร์ วันทิด (วันอาทิตย์) วันพั้ง (วันข้างขึ้น) วันแฮม (วันข้างแรม) เป็นต้น

1.3) คำเรียกเดือน เช่น เดือนหนึ่ง เดือนสอง เดือนสาม เดือนเจียง (เดือนอ้าย) เดือนสิบสอง เป็นต้น

- 2) วิธีเก็บสมุนไพร เช่น บูชา (บูชา) ขุด ฟัน (ตัด) เป็นต้น
- 3) ชื่อพิธีกรรมอื่น ๆ เช่น สะหั่ว (สระผม) ปั้นหน้า (หันหน้า) เป็นต้น
- 4) คำกริยาทั่วไป เช่น เอา ให้ มา ว่า (พูด) จักเอา (จะเอา) ฮัก (รัก) เป็นต้น
- 5) คำนามทั่วไป แบ่งเป็น 2 หมวด ได้แก่

เป็นต้น

5.1) คำนาม/สรรพนาม เช่น หมอยา แม่เขา พ่อเขา ปู่ ญา (ย่า) หมอยา

5.2) ชื่อทิส เช่น ก้าเหนื่อ (ทิสเหนื่อ) ก้าใต้ (ทิสใต้) เป็นต้น

5.3) ลักษณะนาม

5.4) คำนามอื่น

- 6) คำวิเศษณ์ แบ่งเป็น 2 หมวด ได้แก่

6.1) คำวิเศษณ์หน้ากริยา เช่น ลี (จะ) บ่ (ไม่) ลี (ค่อนข้าง) เป็นต้น

6.2) คำวิเศษณ์หลังกริยา เช่น ทด ๆ นำ (ด้วย) เป็นต้น

- 7) คำคุณศัพท์ แบ่งเป็น 2 หมวด ได้แก่

7.1) บอกกำหนด (Determining Adjectives) เช่น นี้ นั้น ไค คำอิด (แรก) ลุน (หลัง) ก่อน ก่อน ๆ เป็นต้น

7.2) บอกคุณลักษณะ คำแดง ขาว เหลือง จ้อย (หอม) เหลือง เป็นต้น

- 8) คำสันธาน เช่น กับ (กับ, และ) กือ ก็ ก็ดี เป็นต้น

9) คำบุพบท เช่น นำ (ตาม) กาง (กลาง) ไว้ ข้าง เเทิง (บน, เหนือ) ใต้ (ล่าง) แต่ (จาก) ญ้อน (เพราะ) คั่นว่า (หากว่า) คั่น (หาก) เป็นต้น

- 10) คำลงท้ายประโยค เช่น แล ดีแล เทอญ เป็นต้น

อนุกรมวิธานคำศัพท์ในตำรายาโบราณอีสาน		
1. ตำรับยา		2. อื่น ๆ
1.1 ยา	1.5 วิธีใช้ยา	2.1 วันเวลา
1.1.1 ชื่อยารักษามนุษย์	1.5.1 วิธีใช้ยา	2.1.1 คำเรียกเวลา
1.1.2 ชื่อยารักษาสัตว์	1.5.2 ข้อห้ามในการใช้ยา	2.1.2 คำเรียกวัน
1.2 อาการป่วย	1.6 สรรพคุณยา	2.1.3 คำเรียกเดือน
1.3 สมุนไพร	1.7 คำรักษา	2.2 วิธีเก็บสมุนไพร
1.3.1 พืชวัตถุ	1.8 คาถา	2.3 ชื่อพิธีกรรมอื่น ๆ
1) ชื่อพืชวัตถุ	1.8.1 คาถารักษาโรค	2.4 คำกริยาทั่วไป
2) ชื่อส่วนประกอบของพืชวัตถุ	1.8.2 คาถาเสกใส่สมุนไพร	2.5 คำนามทั่วไป
3) ลักษณะนามพืชวัตถุ	1.8.3 คาถาเก็บสมุนไพร	2.5.1 คำนาม/สรรพนาม
1.3.2 สัตว์วัตถุ	1.9 ผู้ป่วย	2.5.2 ชื่อทิส
1) ชื่อสัตว์วัตถุ	1.9.1 มนุษย์	2.5.3 ลักษณะนาม
2) ชื่ออวัยวะสัตว์วัตถุ	1) ชื่อ/สรรพนามมนุษย์	2.5.4 คำนามอื่น ๆ
3) ลักษณะนามสัตว์วัตถุ	2) คำเรียกเพศมนุษย์	2.6 คำวิเศษณ์
1.3.3 ธาตุวัตถุ	3) ชื่ออวัยวะมนุษย์	2.6.1 คำวิเศษณ์หน้ากริยา
1) ชื่อธาตุวัตถุ	1.9.2 สัตว์	2.6.2 คำวิเศษณ์หลังกริยา
2) ลักษณะนามธาตุวัตถุ	1) ชื่อสัตว์/สรรพนามสัตว์	2.7 คำคุณศัพท์
1.3.4 กระจกยา	2) คำเรียกเพศสัตว์	2.7.1 บอกกำหนด
1.4 วิธีเตรียมยา/ปรุงยา	3) ชื่ออวัยวะสัตว์	2.7.2 บอกคุณลักษณะ
1.4.1 วิธีเตรียมยา/ปรุงยา	1.10 พิธีกรรมรักษาโรค	2.8 คำสันธาน
1.4.2 จำนวน/ปริมาณ	1.11.1 เครื่องประกอบพิธีกรรม	2.9 คำบุพบท
1.4.3 หน่วยวัดและชั่งตวง	1.11.2 พิธีกรรม	2.10 คำลงท้ายประโยค
1.4.4 อุปกรณ์ปรุงยา		
1.4.5 รูปแบบผลิตภัณฑ์		

รูปที่ 4.20 อนุกรมวิธานคำศัพท์ในตำรายาโบราณอีสาน

4.3 การศึกษาลักษณะอักษรไทยน้อย และอักษรวิธีอักษรไทยน้อยที่ใช้บันทึกตำรายา

โบราณอีสานในใบลาน

การศึกษาครั้งนี้ รวบรวมและวิเคราะห์ข้อมูลจากสำเนาภาพดิจิทัลของตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย โดยเป็นภาพจากใบลานต้นฉบับของตำรายาวิกรมหาชัย จังหวัดมหาสารคาม เล่ม 5 และ ตำรายาวิกรมท่าม่วง อำเภอเสลภูมิ จังหวัดร้อยเอ็ด เล่ม 2 จำนวนทั้งสิ้น 459 ภาพ จำแนกผลการวิจัยเป็น 3 ด้าน ได้แก่ รูปอักษรไทยน้อย ภาษา และอักษรวิธี ดังนี้

4.3.1 รูปอักษรไทยน้อย

รูปอักษรไทยน้อย แบ่งเป็น 4 ประเภท ได้แก่ พยัญชนะ สระ ตัวเลข และเครื่องหมายอื่น ๆ และพบว่า ไม่พบการใช้รูปวรรณยุกต์ในการเขียน ดังนี้

4.3.1.1 พยัญชนะ แบ่งเป็น 4 ประเภท ได้แก่ พยัญชนะเดี่ยว และพยัญชนะประสม พยัญชนะควบกล้ำ และพยัญชนะตัวสะกด รายละเอียดมีดังนี้

1) พยัญชนะเดี่ยว ใช้สำหรับเป็นพยัญชนะต้นและพยัญชนะตัวสะกดรวมทั้งสิ้น 26 รูป โดยมี 1 รูป คือ หอยหยาดน้ำ ที่ในภาษาไทยมาตรฐานจะใช้รูปพยัญชนะ อย และอักษรไทยน้อยที่ไม่พบว่าถูกนำไปเขียนในตำรายามี 18 ตัว ได้แก่ ค ฅ ฉ ฆ ฌ ญ ฎ ฏ ฐ ฑ ฒ ฌ ษ ฌ ทร พ ฤ และ ฤ ตัวอย่างรูปพยัญชนะเดี่ยวอักษรไทยน้อยที่เขียนเป็นพยัญชนะต้นในตำรายาโบราณอีสาน ดังแสดงในรูปที่ 4.21

ก	ข	ค	ง	จ	ช	ด	ต	ถ
ท	น	บ	ป	ผ	ฝ	พ	ฟ	ม
ย	ล	ว	ส	ห	อ	ฮ	หอยหยาดน้ำ	

รูปที่ 4.21 ตัวอย่างรูปพยัญชนะอักษรไทยน้อยที่ใช้เขียนเป็นพยัญชนะต้น

พยัญชนะตัวสะกดอักษรไทยน้อย แบ่งตามมาตราตัวสะกดของอักษรไทยปัจจุบัน 8 มาตรา ได้แก่ แม่กก กค กบ กน กก กม เกย และ เกอว มีมาตราละ 1 รูป 1 เสียง ตัวอย่างคำศัพท์อักษรไทยน้อยที่มีตัวสะกดแบ่งตามมาตราตัวสะกดดังแสดงในตารางที่ 4.6

ตารางที่ 4.6 รูปพยัญชนะตัวสะกดอักษรไทยน้อยและตัวอย่างคำศัพท์แบ่งตามมาตราตัวสะกด

มาตรา	พยัญชนะตัวสะกด (อักษรไทยน้อย)	ตัวอย่างคำศัพท์ (อักษรไทยน้อย)	คำอ่าน (ภาษาไทยถิ่นอีสาน)
1. แม่ กก	๕ (ก)	๕๕๕๕๕	ผักหวาน
		๕๕๕	ปาก
2. แม่ กค	๕ (ค)	๕๕๕๕๕	ประมาท
		๕๕๕	พืช
3. แม่ กบ	๕ (บ)	๕๕๕๕๕	สิบบาท
4. แม่ กน	๕ (น)	๕๕๕	กิน
		๕๕๕๕๕	ฮากตาล
		๕๕๕๕๕	สิทธิการ
		๕๕๕๕๕	ประมาณ
5. แม่ กง	๕ (ง)	๕๕๕๕๕	เทอญ
		๕๕๕๕๕	ยางคิบ
6. แม่ กม	๕ (ม)	๕๕๕๕๕	เห่าห้อม
7. แม่ เกย	๕ (ย)	๕๕๕๕	หน่วย
8. แม่ เกอว	๕ (ว)	๕๕๕	ขาว

2) พยัญชนะประสม ได้แก่ พยัญชนะ 2 ตัวที่เขียนควบติดกัน พยัญชนะตัวแรกเป็นรูปตัวเต็มและตัวที่สองเป็นแบบครึ่งตัว พยัญชนะประสมที่พบเขียนในตำรายาโบราณอีสานมี 4 รูป ได้แก่ “หม” “หย” “हन” “สน” โดยรูปพยัญชนะประสม “หม” และ “हन” จะเขียนคล้ายคลึงกันเป็นอย่างมาก ดังนั้น การอ่านให้ถูกต้องจะต้องดูบริบทร่วมด้วย ตัวอย่างการประสมอักษรเป็นคำด้วยพยัญชนะประสมในอักษรไทยน้อย แสดงในตารางที่ 4.7

ตารางที่ 4.7 รูปพยัญชนะประสมของอักษรไทยน้อยและตัวอย่างคำศัพท์

พยัญชนะประสม (อักษรไทยปัจจุบัน)	พยัญชนะประสม (อักษรไทยน้อย)	ตัวอย่างคำศัพท์ (อักษรไทยน้อย)	คำอ่าน (ภาษาไทยถิ่นอีสาน)
หม	๒๑, ๒๒	ꨀꨂꨀꨂꨀꨂ	แก่นหมาก
หน	๒๓, ๒๔	ꨀꨂꨀꨂꨀꨂ	ฝักหนอก
หย	๒๕, ๒๖	ꨀꨂꨀꨂꨀꨂꨀꨂ	หญาผากควาย
สน	๒๗, ๒๘	ꨀꨂꨀꨂ	สน้อย

3) พยัญชนะควบกล้ำ มี 2 รูป ได้แก่ “กล” และ “พล” ในตำราขาโบราณอีสานที่นำมาศึกษา พบคำศัพท์ที่ประสมด้วยพยัญชนะควบกล้ำพบเพียง 2 คำ เท่านั้น คือ “กล่าว” และ “ปล้น” ตัวอย่างการใช้พยัญชนะควบกล้ำมาประสมเป็นคำ ดังแสดงในตารางที่ 4.8

ตารางที่ 4.8 รูปพยัญชนะควบกล้ำในอักษรไทยน้อยและตัวอย่างคำศัพท์

พยัญชนะควบกล้ำ (อักษรไทยปัจจุบัน)	พยัญชนะควบกล้ำ (อักษรไทยน้อย)	ตัวอย่างคำศัพท์ (อักษรไทยน้อย)	คำอ่าน (ภาษาไทยถิ่นอีสาน)
กล	๓๑, ๓๒	ꨀꨂꨀꨂꨀꨂ	จักกล่าว
พล	๓๓	ꨀꨂꨀꨂꨀꨂ	ปล้น

4.3.1.2 ตัวเลข มี 10 รูป ใช้บอกสัดส่วน ปริมาณ หรือ จำนวน สมุนไพร รูปตัวเลขในอักษรไทยน้อย ดังแสดงในรูปที่ 4.22

๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	ตัวเลขอักษรไทยน้อย
0	๑	๒	๓	๔	๕	๖	๗	๘	๙	ตัวเลขอักษรไทยปัจจุบัน

รูปที่ 4.22 ตัวอย่างรูปตัวเลขในอักษรไทยน้อยในตำราขาโบราณอีสาน

4.3.1.3 สระ มี 22 รูป วิธีใช้มีทั้งเหมือนและต่างกับสระในภาษาไทยมาตรฐาน สระอักษรไทยน้อยที่มีวิธีใช้ต่างกับสระในภาษาไทยมาตรฐาน ดังนี้

1) ประสมอักษรเป็นคำด้วยสระออ มี 3 รูปแบบ ได้แก่ คำที่ประสมด้วยสระออมีพยัญชนะ ย เป็นตัวสะกดจะใช้รูป “สระออย” คำที่ประสมด้วยสระออแบบไม่มีตัวสะกดใช้ นิคหิต (◌̣) และคำที่ประสมด้วยสระออกับพยัญชนะตัวสะกดอื่น ๆ ใช้รูปสระออคล้ายกับภาษาไทยมาตรฐาน

2) ประสมอักษรเป็นคำด้วยสระอา หรือ สระเอะ ใช้ ไม้อัน หรือ ไม้ซัด โดยหากเป็นคำที่มีพยัญชนะตัวสะกดแม่ก (ก) พบว่า บางครั้งจะไม่เขียนพยัญชนะตัวสะกด “ก”

3) สระเอีย มี 2 รูป คือ รูปสระเอียสำหรับประสมอักษรเป็นคำแบบมีตัวสะกด (ใช้รูปพยัญชนะ ย ของอักษรธรรมอีสาน) และแบบไม่มีตัวสะกด (ใช้สระเอียของอักษรไทยน้อย)

นอกจากนี้ ยังพบการประสมอักษรในคำเดียวกันด้วยสระที่มีเสียงคล้ายกัน (สระเสียงสั้นและยาว) ได้แก่ สระอิ/อี สระเอือ/เอือ สำหรับคำที่ออกเสียงสระอุทุกคำจะใช้รูปสระอุเขียน และรูปสระที่ไม่พบการนำมาใช้ประสมคำ คือ สระเอาะ เอียะ เอือะ อัวะ และสระโอะไม่ม้วน ตัวอย่างรูปสระอักษรไทยน้อยและการใช้ประสมอักษรเป็นคำ ดังแสดงในตารางที่ 4.9

ตารางที่ 4.9 ตัวอย่างรูปสระอักษรไทยน้อยและตัวอย่างการใช้

สระ (อักษรไทยปัจจุบัน)	สระ (อักษรไทยน้อย)	ตัวอย่างการประสมอักษรเป็นคำ (ภาษาไทยถิ่นอีสาน)	
-ะ	-:	ຮຸ້ງ ຊະເລ (ถิ่นชะเล)	ກະເລີ (กะเลา)
-า	->	ຮາກ (ฮาก)	ຫາ (ทา)
-อิ, -อี	ຸ ອ	ຟຣີ (ฟริก)	ຈີ (จิง)
-อี, -เอี	ື ອ	ຊີ້ (จิ้น)	ຕ່າງ ຈີ້ (ตำลึง)
-อุ, -เอุ	ຸ ອ	ຜູ້ (ฟู)	ຕຸ້ມ (ตุ้ม)
เ-, เ-ะ	เ- >	ແກ້ (แกก)	ແກ້ວ (แก็ด)
แ-ะ	แ-:	ແປ້ (แจะ)	ແປ້ງ (แจะ)
แ-	แ&	ແປງ (แบง)	ແປ້ (แถ)
โ-ะ	โ-:	ຮີ ໂຈ້ (หีโทะ)	ໂພ້ (โพะ)
โ-	โ&	ໂຄຍ ທາ (โคยตาล)	ໂອມ (โอม)
ออ	-อ	ຂອດ (ขอด)	ດອກ (ดอก)
เ-อะ	เ-:	ຊະເອ້ (ชะเออะ)	ຝາເອ້ (ยาเซอะ)
เ-อ เอือ	เ- ^๐ เ- ^๑ ๑	ໜາກ-ຮີ (หมากเว่อ)	ເຄືອ (เครือ)
เ-ีย	เ- ^๑ ๓ -๓	ເສຍ (เสีย)	ເສຍ (เสีย)

ตารางที่ 4.9 ตัวอย่างรูปสระอักษรไทยน้อยและตัวอย่างการใช้ (ต่อ)

สระ (อักษรไทยปัจจุบัน)	สระ (อักษรไทยน้อย)	ตัวอย่างการประสมอักษรเป็นคำ (ภาษาไทยถิ่นอีสาน)	
-ัว	-วิ	๑๖ (ถั่ว)	๗๖ (หัว)
-ัว	-วิ	๖๖ (น้ำน้ำ)	๗๖ (ตำตำ)
ไ-ไ-	๖-	๖๗ (ไหให้)	๖๗ (ใส/ใส่/ใส่)
เ-า	๖(๖)	๖๖(๖) (เขาเข้า/เข้า)	๗๖ (หา)
ออย	๖	๖(๖) (น้อย)	๖๖ (อ้อยคำ)
นิกhit	๖	๖ (บ)	๖ (ห่อ)
ไม้อัน/ซัด	๖	๖๖ (เป็น), ๖๖ (ฮาก)	๖๖ (อัน)
ไม้กง/กุ่ม	๖	๖๖ (ต้ม)	๖๖ (ฝน)

4.3.1.5 เครื่องหมายอื่น ๆ ได้แก่ เครื่องหมายจบคำรับขาแต่ละขนาน มีรูปแบบเฉพาะแตกต่างกันในตำรายาโบราณอีสานแต่ละฉบับ พบการใช้บ่อย ตัวอย่างดังรูปที่ 4.23

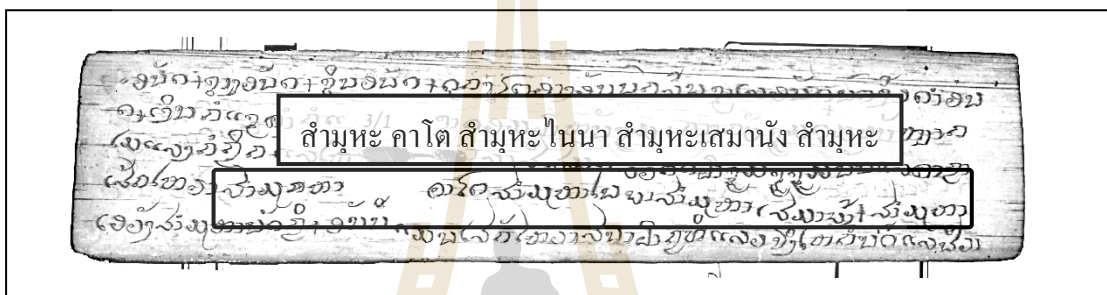


รูปที่ 4.23 ตัวอย่างเครื่องหมายจบคำรับขาโบราณอีสาน

4.3.2 ภาษา

4.3.2.1 ภาษาที่ใช้บันทึกในตำรายาโบราณอีสาน

การบันทึกตำรายาโบราณอีสานด้วยอักษรไทยน้อยบันทึกเป็น 2 ภาษา ได้แก่ ภาษาไทยถิ่นอีสานโบราณ และภาษาบาลี โดยภาษาบาลีจะใช้เขียนคาถาแทรกในเนื้อหาส่วนต่าง ๆ เช่น การไปเก็บสมุนไพร การเตรียมยา/ปรุงยา การใช้ยา หรือรักษาโรคแบบไม่ใช้สมุนไพร หรือใช้ในพิธีกรรม ตัวอย่างดังรูปที่ 4.24



รูปที่ 4.24 ตัวอย่างคาถาภาษาบาลีสำหรับเสกใส่ยา

4.3.2.2 ลักษณะคำศัพท์ในภาษาไทยถิ่นอีสาน

คำศัพท์ภาษาไทยถิ่นอีสานที่ใช้เขียนในตำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยมีคำพ้องเสียง คำพ้องเสียง คือ คำที่ออกเสียงเหมือนกัน แต่มีความหมายต่างกัน คำศัพท์ที่พบในตำรายาโบราณอีสานในการศึกษาครั้งนี้ มีคำพ้องเสียงใน 2 ลักษณะ คือ คำพ้องเสียงภายในภาษาถิ่นอีสานเอง และ คำพ้องเสียงกับภาษาถิ่นอื่น จึงมักทำให้ผู้ที่ไม่รู้จักภาษาไทยถิ่นอีสานเป็นอย่างดีอาจแปลความหมายผิดได้ ตัวอย่างในตารางที่ 4.10

ตารางที่ 4.10 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่พ้องเสียงกับคำศัพท์ภาษาไทยมาตรฐาน

คำศัพท์ (ภาษาไทยถิ่นอีสาน)	ความหมาย (ภาษาไทยถิ่นอีสาน)	ความหมาย (ภาษาไทยมาตรฐาน)
ดั่ง	จุมก (นาม)	บังเกิดเสียงขึ้นหรือทำให้เสียงบังเกิดขึ้นอย่างแรง, มีชื่อเสียง (วิเศษณ์)
พุด	1. ของที่จัดไว้เป็นส่วน ๆ (นาม) 2. ฟั่น (กริยา)	เปล่งเสียงออกเป็นถ้อยคำ (กริยา)
ชวย	กรวยใบตองกล้วย (นาม)	เคราะห์ร้าย, อัปโชค (วิเศษณ์)

ตารางที่ 4.10 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่พ้องเสียงกับคำศัพท์ภาษาไทยมาตรฐาน (ต่อ)

คำศัพท์ (ภาษาไทยถิ่นอีสาน)	ความหมาย (ภาษาไทยถิ่นอีสาน)	ความหมาย (ภาษาไทยมาตรฐาน)
คาย	1. ทำให้หลุดออกมา (กริยา) 2. เครื่องบูชาคุณ (นาม) 3. ระคายเคือง (วิเศษณ์)	1. ปล่อยของที่ไม่ต้องการจะกลืนหรือกลืนไม่ได้ออกจากปาก, ปล่อยออกมา (กริยา) 2. ส่วนที่มีลักษณะเป็นผงหรือขนละเอียดแหลมคม ของบางสิ่งบางอย่างเวลากระทบผิวหนังทำให้รู้สึกระคายคัน เช่น คายข้าว (นาม)

4.3.3 อักษรวิธีการประสมอักษรไทยน้อยเป็นคำ

การประสมอักษรไทยน้อยเป็นคำ ตำแหน่งการเขียนตัวอักษรไทยน้อยมี 3 ระดับ ได้แก่ ระดับสระบน ระดับพยัญชนะ และระดับสระล่าง คล้ายคลึงกับวิธีการเขียนอักษรภาษาไทยปัจจุบัน แต่สิ่งที่แตกต่างคือการเขียนพยัญชนะควบกล้ำหรือพยัญชนะตัวตามบางตัวจะใช้วิธีเขียนของอักษรธรรมอีสาน ได้แก่ คำที่ประสมด้วยพยัญชนะควบกล้ำ “ล” หรือ คำที่มีพยัญชนะต้นที่ตามด้วยพยัญชนะ “ล” จะใช้รูปพยัญชนะตัวสะกด “ล” ของอักษรธรรมอีสานมาเขียนไว้ได้พยัญชนะต้นอักษรไทยน้อย ตัวอย่างดังรูปที่ 4.25

(1) ระดับสระบน	1	อู๋	งั่ม	ฏาว	ฏอติบ	
(2) ระดับพยัญชนะ	2	อู๋	งั่ม	ฏาว	ฏอติบ	
(3) ระดับสระล่าง	3	อู๋	งั่ม	ฏาว	ฏอติบ	
		อู๋	งั่ม	ฏาว	ฏอติบ	(คำปริวรรต)
		(อู๋ดั่งนั้น)	(งั่ม)	(กล่าว)	(ก้วยติบ/ก้วยติบ)	(คำอ่านภาษาไทยถิ่นอีสาน)

รูปที่ 4.25 การประสมอักษรไทยน้อยเป็นคำ

4.3.4 ลักษณะการสะกดคำของอักษรไทยน้อย

4.3.4.1 ไม่นิยมประสมอักษรเป็นคำด้วยพยัญชนะควบกล้ำและตัวการันต์

การบันทึกคำราชาโบราณอีสานพบการประสมอักษรเป็นคำโดยใช้พยัญชนะควบกล้ำค่อนข้างน้อย โดยพยัญชนะควบกล้ำที่พบ ได้แก่ “ล” พบเพียง 2 คำ เท่านั้น ได้แก่ “กล่าว” “พลัน” และพยัญชนะควบกล้ำ “ว” พบการใช้ไม่สม่ำเสมอ สำหรับตัวการันต์ไม่ใช้เลย ตัวอย่างคำศัพท์ที่สะกดด้วยอักษรไทยน้อยที่ไม่ใช้พยัญชนะควบกล้ำและตัวการันต์ ดังแสดงในตารางที่ 4.11

ตารางที่ 4.11 ตัวอย่างคำศัพท์ที่สะกดด้วยอักษรไทยน้อยที่ไม่ใช้พยัญชนะควบกล้ำและตัวการันต์

คำศัพท์อักษรไทยน้อย	คำปริวรรต	คำอ่านภาษาไทยถิ่นอีสาน
ສົງຄວນ	สอยสงวน	สร้อยสังวาล
ພົມເຂັດ	พิมเข็ด	พรหมเชษฐ
ປະໝາດ	ปะหมัด	ประมาท
ຄວຍສຸກ	กวยสุก	กล้วยสุก
ລູກຄອນ	ลูกกอน	ลูกกลอน
ເຂົາກວາງ	เขากวาง	เขากวาง
ຄວຍດ່ອນ	กวยด่อน	กล้วยด่อน
ຟັກ	ฟัก	พริก/พลิก

4.3.4.2 คำเดียวกันสะกดคำได้หลายแบบ

การสะกดคำด้วยอักษรไทยน้อย คำเดียวกันสามารถสะกดได้หลายแบบ เนื่องจากไม่มีหลักเกณฑ์ในการสะกดคำที่เป็นมาตรฐานชัดเจน ต่างจากภาษาไทยปัจจุบันที่ใช้หลักเกณฑ์การสะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน โดยสรุปลักษณะการสะกดคำได้ดังนี้

1) ใช้พยัญชนะต้นที่มีหน่วยเสียงคล้ายกันเขียนสะกดคำเดียวกัน ได้แก่ ใช้พยัญชนะ “จ/ค” เขียนสะกดคำที่มีเสียง “จ/ค/ฅ” ใช้พยัญชนะ “ซ/ส” เขียนสะกดคำที่มีเสียง “ฌ/ซ/ส” ใช้พยัญชนะ “ฟ/ฝ” เขียนสะกดคำที่มีเสียง “ฟ/ฝ” และใช้พยัญชนะ “ล” เขียนสะกดคำที่มีเสียง “ร/ล” ตัวอย่างแสดงดังรูปที่ 4.26

ฉิว	ฉิว	โศก	โศก	สะบับ	ชะบับ	คำอักษรไทยน้อย
ฟัน	ฟัน	โคขาว	โงขาว	สะบับ	ชะบับ	คำปริวรรต
ฝน		โงขาว		ลบบับ		ภาษาไทยถิ่นอีสาน

รูปที่ 4.26 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อย กรณีคำเดียวกันแต่ใช้พยัญชนะต้นต่างกัน

2) ใช้สระเสียงสั้นและสระเสียงยาวเขียนสะกดคำเดียวกัน ได้แก่ สระ อี/อี สระอิ/อิ สระเอือ/เอือ และ สระอะ/อา ตัวอย่างดังแสดงในตารางที่ 4.12

ตารางที่ 4.12 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อย กรณีคำเดียวกันใช้ทั้งสระเสียงสั้นและสระเสียงยาวในการสะกด

สระ (อักษรไทยน้อย)	สระ (อักษรไทย)	คำศัพท์ (อักษรไทยน้อย)	คำปริวรรต	คำอ่าน (ภาษาไทยถิ่นอีสาน)
อ, อ	ิ, ี	อิ, ี	ผี	ผี, ผี
อ	ิ, ี	เอือเอือ	เขือเขิน	เขือเขิน
		เอือเอือเอือ	เขามีนเขิน	เขามีนเขิน
อ, อ	เอ, เอือ	เอือ, เอือ	เหือด	เหือด

3) ใช้อักษรธรรมอีสานมาประสมอักษรเป็นคำร่วมกับอักษรไทยน้อย โดยใช้พยัญชนะรูปตัวเชิงของอักษรธรรมอีสานที่มีเสียงตรงกับพยัญชนะของอักษรไทยน้อย มาเป็นตัวสะกด หรือตัวตาม ได้แก่ อักษร ส ข น บ และ พ ดังตัวอย่างในตารางที่ 4.13

ตารางที่ 4.13 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อยผสมกับอักษรธรรมอีสาน

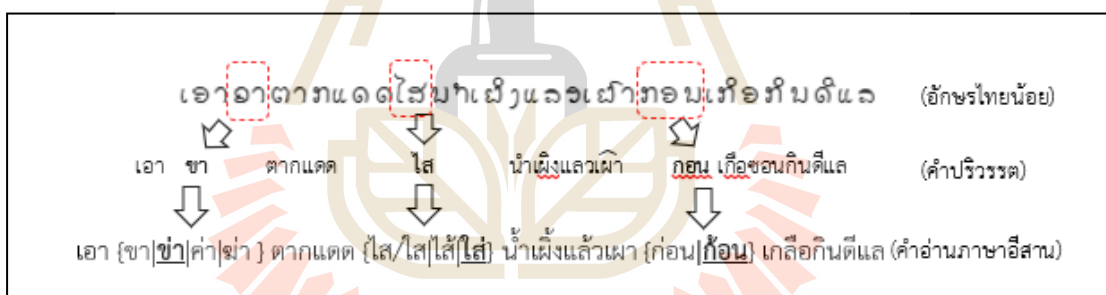
พยัญชนะ (อักษรธรรมอีสาน)	คำศัพท์ (อักษรไทยน้อย)	คำปริวรรต	คำอ่าน (ภาษาไทยถิ่นอีสาน)
ส (ย)	สั้ง	พิษ	พิษ
	สั้ง	พันชาย	พันชาด
	สั้ง	บาช	บาด/บาท
	สั้ง	ผักกาศ	ผักกาด
ส (ย)	สั้ง-บั้ง	สาละบาส	สารบาท

ตารางที่ 4.13 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อยผสมกับอักษรธรรมอีสาน (ต่อ)

พยัญชนะ (อักษรธรรมอีสาน)	คำศัพท์ (อักษรไทยน้อย)	คำปริวรรต	คำอ่าน (ภาษาไทยถิ่นอีสาน)
๓ (ส)	๓๓๓๓	กัตสปะ	กัตสปะ
	๓๓	กับ	กับ
๓ (พ)	๓๓๓	สัพพา	สัพพา
๓ (น)	๓๓	ไบลาน	ไบลาน

4.3.4.3 สะกดคำโดยไม่มีวรรณยุกต์และอ่านได้หลายแบบ

อักษรไทยน้อยไม่มีวรรณยุกต์บังคับการออกเสียง เมื่อนำมาบันทึกภาษาไทยถิ่นอีสานซึ่งเป็นภาษาที่มีโทนเสียงวรรณยุกต์ การอ่านคำ ผู้อ่านจึงต้องผันเสียงให้ตรงกับ ความหมายหรือบริบทของเรื่องที่อ่านเอาเอง ซึ่งจากตัวอย่างการสะกดคำด้วยอักษรไทยน้อยในรูปที่ 4.27 จะเห็นได้ว่าคำศัพท์ในตำรายาโบราณอีสานบางคำมีคำอ่านมากกว่า 1 คำอ่าน และการจะอ่านให้ถูกต้องจึงจำเป็นต้องมีความรู้ทั้งภาษาไทยถิ่นอีสานโบราณ และความรู้เรื่องตำรายาอีสาน



รูปที่ 4.27 ตัวอย่างการสะกดคำด้วยอักษรไทยน้อยที่ไม่มีวรรณยุกต์

4.3.4.4 ตัวเลขของอักษรไทยน้อยบางรูปคล้ายคลึง หรือมีรูปเดียวกับพยัญชนะ และใช้เขียนแทนคำศัพท์ได้

ตัวเลขของอักษรไทยน้อยที่มีรูปคล้ายคลึง หรือมีรูปเดียวกับพยัญชนะที่ พบในตำรายาโบราณอีสาน ได้แก่ เลข “1” คล้ายกับพยัญชนะ “ด” เลข “6” คล้ายกับพยัญชนะ หอย หนาดน้ำ (ย-หางยาว) และเลข “7” มีรูปเดียวกับพยัญชนะ “ถ” และตัวเลขที่มักนำมาใช้เขียนแทนคำคือ เลข 2 โดยใช้แทนการเขียนคำเดิมซ้ำครั้งที่ 2 (เหมือนเขียนไม้ยมกในอักษรไทยปัจจุบัน) หรือเขียน แทนคำที่อ่านออกเสียงว่า “สอง” ซึ่งทำให้เกิดความสับสนในการจำแนกว่าตัวเลขที่เขียนนั้นเป็น ตัวเลขแสดงปริมาณ เป็นพยัญชนะ หรือการเขียนคำเดิมซ้ำ การปริวรรตตัวเลขอักษรไทยน้อยจึงต้อง ใช้คำบริบทที่อยู่รอบ ๆ มาช่วยให้การระบุตัวอักษรให้ถูกต้อง ตัวอย่างการเขียนตัวเลขดังแสดงใน ตารางที่ 4.14

ตารางที่ 4.14 ตัวอย่างการเขียนตัวเลขของอักษรไทยน้อยในคำราชาโบราณอีสาน

คำศัพท์ (อักษรไทยน้อย)	คำปริวรรต	คำอ่าน (ภาษาไทยถิ่นอีสาน)	ความหมาย
๗	๗	เจ็ด	7
๕๕๐	๕๕๐	ห้าร้อย	500
๑๑๑๑๑	แดง๒	แดง ๆ	ใช้ ๒ แทน ๆ
๒๐	๒๐	๒๐	20
๒๐๐	๒๐๐	สองร้อย	ชื่อโรค/อาการ

4.3.4.5 การสะกดคำด้วยอักษรวิเศษที่ขึ้นอยู่กับผู้เขียนแต่ละคน

การสะกดคำโดยใช้อักษรวิเศษ หรือลูกเล่นในการสะกดคำที่ขึ้นอยู่กับผู้เขียนแต่ละคน ทำให้คำนั้น ๆ อ่านยาก เช่น ตำแหน่งการเขียนสระหรือพยัญชนะต่างจากการเขียนปกติ ใช้พยัญชนะร่วมกันในคำเดียวกัน ละพยัญชนะตัวสะกด ใช้อักษรวิเศษของอักษรธรรมมาเขียนสะกดคำโดยใช้พยัญชนะหรือสระ ไม่ตรงกับเสียงอ่าน หรือรูปอักษรไทยน้อยบางรูปเป็นได้ทั้ง ตัวเลข พยัญชนะ และ เป็นคำ เป็นต้น ตัวอย่างดังแสดงในตารางที่ 4.15

ตารางที่ 4.15 ตัวอย่างการสะกดคำด้วยอักษรวิเศษที่ขึ้นอยู่กับผู้เขียนแต่ละคน

คำศัพท์ (อักษรไทยน้อย)	คำปริวรรต	คำอ่าน (ภาษาไทยถิ่นอีสาน)	คำอธิบาย
๑๑๑	แก	แก	รูปสระแอและตำแหน่งการวางสระต่างจากวิเศษปกติ
๑๑๑	แก	แก	
๑๑๑๑๑	ฮ้า(ก)หม้า(ก)บ	ฮากหมากบ้ำ	ละพยัญชนะตัวสะกด
๑๑๑๑๑	๑๑๑๑๑	๑๑๑๑๑	ใช้พยัญชนะร่วมกันระหว่างคำหรือพยางค์
๑๑๑๑๑	๑๑๑๑๑	๑๑๑๑๑	
๑๑๑๑๑	พระพุททีเจ้า	พระพุททเจ้า	เขียนตัวควบกล้ำ “ร” ไว้หน้าพยัญชนะต้นตามแบบอักษรธรรมอีสาน
๑๑๑	กรม	กรม	รูปอักษรแบบเดียวกันอ่านต่างกัน (“๑” และ “๑” มีรูปเดียวกัน, “๑” และ ๑ มีรูปเดียวกัน)
๑๑๑๑	๑๑๑๑	๑๑๑๑	
๑๑๑๑	๑๑๑๑	๑๑๑๑	
๑๑๑๑	๑๑๑๑	๑๑๑๑	คำอ่านไม่ตรงกับการเขียน
๑๑๑๑	๑๑๑๑	๑๑๑๑	

4.4 ลักษณะอักษรไทยน้อยที่ส่งผลกระทบต่อการใช้บรรทัดคำภาษาไทยโบราณอีสานแบบอัตโนมัติ

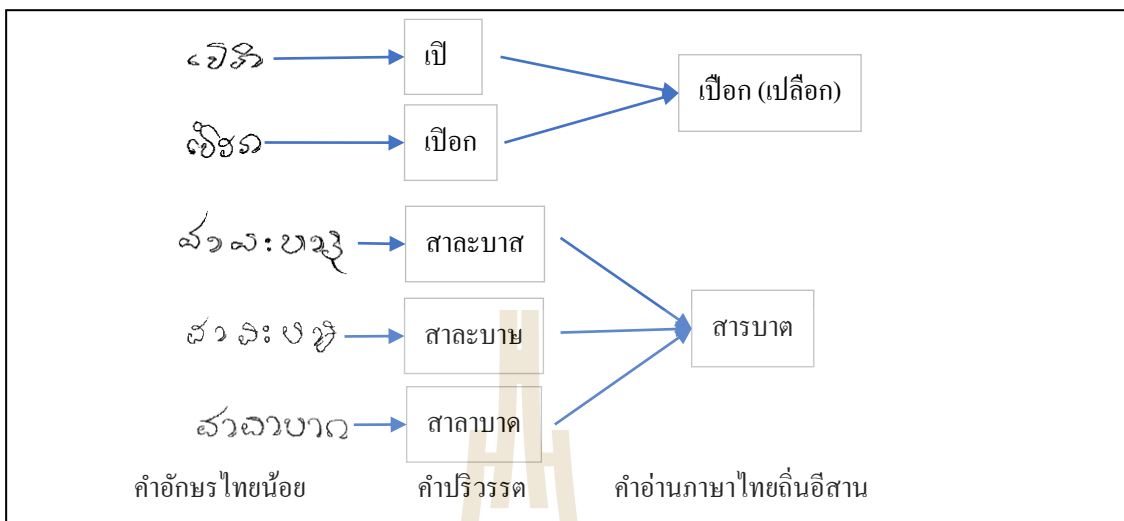
จากผลการศึกษาลักษณะอักษรไทยน้อยที่ใช้บันทึกคำภาษาไทยโบราณอีสานในด้านต่าง ๆ ได้แก่ รูปอักษร ภาษา อักษรวิธีการประสมอักษร และการสะกดคำ ในคำภาษาไทยโบราณอีสาน สรุปผลกระทบที่มีต่อการปริวรรตคำภาษาไทยโบราณอีสานแบบอัตโนมัติ แบ่งเป็น 2 ส่วน ได้แก่ การทำคำปริวรรต และ การทำคำอ่านภาษาไทยถิ่นอีสาน ดังนี้

4.4.1 ปัญหาการทำคำปริวรรต

การปริวรรต (Transliterations) เป็นกระบวนการถ่ายทอด แปลงอักษร ภาษา และ อักษรวิธี การเขียนระบบหนึ่งไปสู่อีกระบบหนึ่งอย่างมีหลักการ ซึ่งในการศึกษาครั้งนี้คือ การถ่ายทอดคำศัพท์ภาษาไทยถิ่นอีสานโบราณที่เขียนด้วยอักษรไทยน้อยให้อยู่ในรูปอักษรไทยปัจจุบัน แต่ยังคงลักษณะการสะกดคำตามภาษาไทยถิ่นอีสานโบราณที่บันทึกไว้ในเอกสารโบราณต้นฉบับ พบปัญหาในการถอดอักษรไทยน้อยที่บันทึกในคำภาษาไทยถิ่นอีสานเป็นอักษรไทยปัจจุบัน ได้แก่ (1) รูปตัวเลขของอักษรไทยน้อยบางรูปมีรูปเดียวกับพยัญชนะ หรือคล้ายคลึงกับพยัญชนะ ได้แก่ เลข “1” กับ “ด” และ เลข “7” กับ “ถ” ดังนั้น การถอดคำปริวรรตให้ถูกต้องจะต้องอาศัยบริบทแวดล้อม (2) การเขียนอักษรธรรมอีสานปะปนกับอักษรไทยน้อย ดังนั้น ผู้ปริวรรตคำภาษาไทยถิ่นอีสานจำเป็นต้องรู้จักทั้งอักษรไทยน้อยและอักษรธรรมอีสาน

4.4.2 ปัญหาการทำคำอ่านภาษาไทยถิ่นอีสาน

การบันทึกภาษาไทยถิ่นอีสานโบราณด้วยอักษรไทยน้อยมีลักษณะพิเศษคือ (1) ไม่มีวรรณยุกต์กำกับ โทณเสียง เช่น คำว่า “ใจ” และ “ใจ” จะเขียนเหมือนกันคือ “ใจ” (2) อักษรบางตัวมีเสียงเหมือนกัน เช่น “ข/ค” “ซ/ส” (3) มีการยืมพยัญชนะของอักษรธรรมอีสานบางตัวมาใช้ ลักษณะเหล่านี้ ทำให้เกิดปัญหาสำคัญคือ “คำหนึ่งคำสะกดได้หลายแบบ” หรือ “มีคำปริวรรตหลายแบบ/หลายคำ” ซึ่งการสะกดคำปริวรรตแต่ละแบบอาจทำให้เกิดความเข้าใจคลาดเคลื่อนและทำอ่านออกมาเป็นคำคนละคำ ทั้งที่จริงคือคำเดียวกัน มีคำอ่านเดียวกัน ดังตัวอย่างรูปที่ 4.29



รูปที่ 4.29 ตัวอย่างคำเดียวกันแต่สะกดคำต่างกัน

ทั้งนี้เนื่องจาก (1) จำนวนพยัญชนะและสระของอักษรไทยน้อยมีจำนวนน้อยกว่าอักษรไทยปัจจุบัน (2) การไม่มีวรรณยุกต์กำกับ โทณเสียง และ (3) การประสมอักษรเป็นคำด้วยอักษรวิธีพิเศษเฉพาะของผู้เขียนแต่ละคน ส่งผลให้คำศัพท์ภาษาไทยถิ่นอีสานที่เขียนด้วยอักษรไทยน้อยหลายคำมีลักษณะ “เขียนอย่างหนึ่งแต่อ่านอีกอย่างหนึ่ง” ขึ้นอยู่กับบริบทและเรื่องที่อ่าน แม้ถ่ายทอดเป็นคำปริวรรตอักษรไทยปัจจุบันแล้ว แต่ยังคงอ่านยาก

การศึกษาครั้งนี้สรุปลักษณะความสัมพันธ์ระหว่างคำปริวรรตและคำอ่านภาษาไทยถิ่นที่พบในตำราโยธาโบราณอีสานแบ่งได้เป็น 4 รูปแบบ ได้แก่

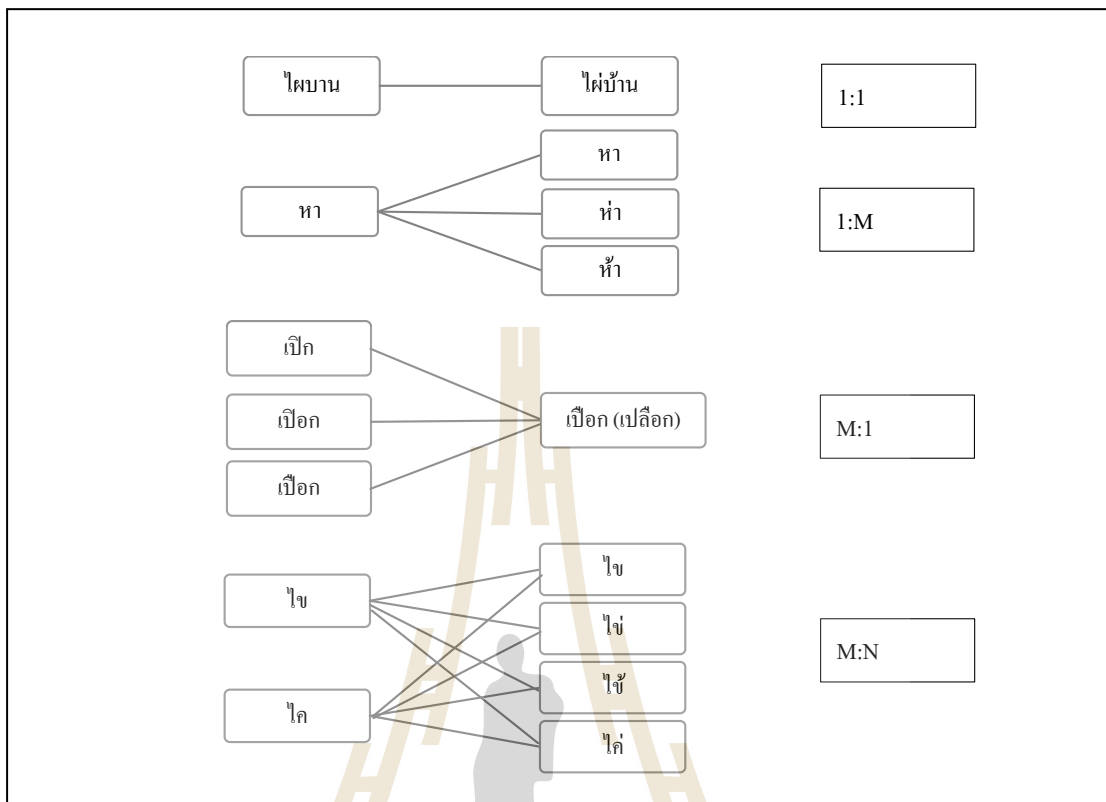
4.4.2.1 แบบ 1:1 หมายถึง คำปริวรรต 1 คำ มีคำอ่านภาษาไทยถิ่นอีสานได้ 1 คำ

4.4.2.2 แบบ 1:M หมายถึง คำปริวรรต 1 คำ มีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำอ่าน

4.4.2.3 แบบ M:1 หมายถึง คำปริวรรตหลายคำมีคำอ่านภาษาไทยถิ่นอีสาน 1 คำอ่าน (อ่านได้แบบเดียว)

4.4.2.4 แบบ M:N หมายถึง คำปริวรรตแต่ละคำมีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำอ่าน และคำอ่านภาษาไทยถิ่นอีสานแต่ละคำมีคำปริวรรตได้หลายคำ

แสดงตัวอย่างดังในรูปที่ 4.30



รูปที่ 4.30 ตัวอย่างคำปริวรรตและคำอ่านภาษาไทยถิ่นอีสานจำแนกตามประเภทความสัมพันธ์

จากรูปที่ 4.30 จะเห็นว่า ความสัมพันธ์ระหว่างคำปริวรรตและคำอ่านภาษาไทยถิ่นอีสาน กรณีความสัมพันธ์แบบ 1:1 และแบบ M:1 เป็นรูปแบบที่คำปริวรรต 1 คำ หรือหลายคำ มีคำอ่านได้เพียงคำอ่านเดียว ซึ่ง 2 กรณีแรกนี้ สามารถแก้ปัญหาการแปลงคำปริวรรตเป็นคำอ่านแบบอัตโนมัติได้โดยการใช้คลังศัพท์ (Lexicon) หรือพจนานุกรม ที่มีการเชื่อมโยงระหว่างคำศัพท์หลัก (Descriptor) คือ “คำปริวรรต” กับ “คำอ่าน” ที่เป็นคำพ้องความหมาย (Synonyms) และระหว่างคำศัพท์หลักกับ “คำอ่าน” ที่เป็นคำพ้องเสียง (Homonyms) หรือ คำพ้องรูป (Homograph) ในขณะที่กรณีที่ 3 และ 4 ที่เป็นความสัมพันธ์แบบ 1:M และ M:N ที่คำปริวรรต 1 คำ มีคำอ่านได้มากกว่า 1 คำ ไม่สามารถแก้ปัญหาการแปลงคำปริวรรตเป็นคำอ่านที่ถูกต้องได้โดยใช้คลังศัพท์เพียงอย่างเดียว แต่ต้องอาศัยบริบทของคำที่อยู่รอบข้างเข้ามาช่วย โดยใช้ ออนโทโลยี (Ontology) เป็นเครื่องมือแสดงความสัมพันธ์ที่ซับซ้อนระหว่างคำศัพท์แต่ละกลุ่ม ดังนั้น ในการพัฒนารอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยให้เป็นคำอ่านภาษาไทยถิ่นอีสานในครั้งนี้ จึงได้พัฒนากล้องศัพท์ด้านคำราชาภาษาไทยถิ่นอีสาน และออนโทโลยีด้านคำราชาอีสานขึ้นมา เพื่อใช้เป็นเครื่องมือสำหรับการแปลงคำปริวรรตไปเป็นคำอ่าน โดยมีรายละเอียดดังแสดงในบทที่ 5

บทที่ 5

ผลการพัฒนาและประเมินประสิทธิภาพรอบการทำงาน ในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

ผลการพัฒนาและประเมินประสิทธิภาพรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานนี้ เป็นผลการวิจัยระยะที่ 2 และ 3 รายละเอียดผลการวิจัยมีดังนี้

- 5.1 การพัฒนาคล้งศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน
 - 5.1.1 ลักษณะคล้งศัพท์ภาษาไทยถิ่นอีสานสำหรับการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานแบบอัตโนมัติ
 - 5.1.2 การวิเคราะห์และออกแบบโครงสร้างคล้งศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน
 - 5.1.3 เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสาน
 - 5.1.4 การพัฒนาคล้งศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน
- 5.2 การพัฒนาออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน
 - 5.2.1 ความต้องการออนโทโลยีสำหรับแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน
 - 5.2.2 การพัฒนาออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน
 - 5.2.3 การประเมินประสิทธิภาพของออนโทโลยี
- 5.3 การออกแบบขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน
 - 5.3.1 การตัดคำ
 - 5.3.2 การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคล้งศัพท์
 - 5.3.3 การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยออนโทโลยีคำศัพท์
- 5.4 การประเมินประสิทธิภาพรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน
 - 5.4.1 ความถูกต้องแม่นยำ
 - 5.4.2 ความผิดพลาดของตัวแบบ
 - 5.4.2 ประสิทธิภาพการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

5.1 การพัฒนาคำศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

5.1.1 ลักษณะคำศัพท์ภาษาไทยถิ่นอีสานสำหรับการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานแบบอัตโนมัติ

การศึกษาลักษณะคำศัพท์ภาษาไทยถิ่นอีสาน ประกอบด้วย การวิเคราะห์โครงสร้างหรือองค์ประกอบในการอธิบายคำศัพท์ภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐานจากพจนานุกรมรวมทั้งสิ้น 8 เล่ม ได้แก่ (1) สารานุกรมภาษาไทยถิ่นอีสาน-ไทย-อังกฤษ (ปรีชา พิณทอง, 2532) (2) ประมวลคำศัพท์ภาษาลาวในเอกสารโบราณ เล่ม 1 - 5 (หอสมุดแห่งชาติ, 2544) (3) พจนานุกรม 3 ภาษา (ไทยน้อย-อังกฤษ-ไทย) (รัตนา จันทร์เทาว์, 2558) (4) พจนานุกรมภาคอีสาน-ภาคกลาง ฉบับปณิธานสมเด็จพระมหาวิรรวงศ์ (ดิศฺสมหาเถระ) (สมเด็จพระมหาวิรรวงศ์ (ดิศฺสมหาเถระ), 2541) (5) พจนานุกรมภาษาถิ่นอีสาน : เว้าอีสาน (บุญเกิด พิมพ์วรมงคล และ นภาพร พิมพ์วรมงคล, 2545) (6) พจนานุกรมภาษาไทยถิ่นอีสาน ฉบับ คำพูน บุญทวี (วิระ สุตสังข์ และ ลัณนา เจริญสิทธิชัย, 2548) (7) พจนานุกรมภาษาไทยถิ่นอีสาน-ไทยกลาง (สำลี รักสุทธี, 2554) และ (8) พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (ราชบัณฑิตยสถาน, 2557) องค์ประกอบที่ทำการวิเคราะห์มีจำนวน 11 องค์ประกอบ ได้แก่ อักษรที่ใช้เขียน ภาษา หมวดคำศัพท์ มีคำศัพท์อีสานโบราณ มีการแบ่งหมวดคำในคลังศัพท์ (POS) มีอ้างอิงที่มาของคำ มีความหมาย/คำแปล มีตัวอย่างประโยค มีคำตรงข้าม มีคำคล้ายคลึง และมีคำปริวรรต แสดงรายละเอียดผลการวิเคราะห์ดังตารางที่ 5.1

ตารางที่ 5.1 ผลการวิเคราะห์โครงสร้างการอธิบายคำศัพท์ของคลังศัพท์ภาษาไทยมาตรฐานและ
ภาษาไทยถิ่นอีสาน

คลังศัพท์/พจนานุกรม	องค์ประกอบที่ใช้อธิบายคำศัพท์ในคลังศัพท์/พจนานุกรม										
	อักษรที่ใช้พิมพ์คลังศัพท์หลัก	ภาษา	หมวดคำ	มีคำศัพท์อีสานโบราณ	มีชนิดของคำ (POS)	มีที่มาของคำ/อ้างอิง	มีความหมาย/คำแปล	มีตัวอย่างประโยค	มีคำตรงข้าม	มีคำคล้ายคลึง	มีคำปริวรรต
พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554	ไทย	ไทย	ทั่วไป	✓	✓	✓	✓	✓	✓	✓	✗
สารานุกรมภาษาไทยถิ่นอีสาน-ไทย-อังกฤษ	ไทย, อังกฤษ	อีสาน, ไทย, อังกฤษ	ทั่วไป, คำราชาอีสาน*	✓	✓	✓	✓	✓	✗	✗	✗
ประมวลคำศัพท์ภาษาลาวในเอกสารโบราณ เล่ม 1-5	ไทย	ลาว, อีสาน	คำภาษาไทยโบราณ, คำราชาอีสาน*	✓	✓	✓	✓	✗	✗	✗	✗
พจนานุกรม 3 ภาษา (ไทยน้อย-อังกฤษ-ไทย)	ไทยน้อย, ไทย, อังกฤษ	อีสาน, ไทย	ทั่วไป, คำราชาอีสาน*	✓	✓	✗	✓	✗	✗	✗	✗
พจนานุกรมภาคอีสาน-ภาคกลาง ฉบับปณิธาน สมเด็จพระมหาวิริวงศ์ (ติสุสมหาเถระ)	ไทย	ไทย, อีสาน	ทั่วไป, คำราชาอีสาน*	✓	✓	✓	✓	✓	✗	✗	✗
พจนานุกรมภาษาถิ่นอีสาน: เว้าอีสาน	ไทย	ไทย, อีสาน	ทั่วไป, คำราชาอีสาน*	✓	✓	✗	✓	✗	✗	✗	✗
พจนานุกรมภาษาไทยถิ่นอีสาน ฉบับ คำพูน บุญทวี	ไทย	ไทย, อีสาน	ทั่วไป	✓	✗	✗	✓	✗	✗	✗	✗
พจนานุกรมภาษาไทยถิ่นอีสาน-ไทยกลาง	ไทย	ไทย, อีสาน	ทั่วไป, คำราชา*	✓	✓	✗	✓	✓	✗	✓	✗
คลังศัพท์สำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน	ไทย	ไทย, อีสาน	ทั่วไป, คำราชาโบราณอีสาน	✓	✓	✓	✓	-	-	✓	✓

ผลการวิเคราะห์โครงสร้างหรือองค์ประกอบในการอธิบายคำศัพท์ของคลังศัพท์ภาษาไทยถิ่นอีสานที่มีอยู่ในปัจจุบันพบว่า มีข้อจำกัดในการนำไปใช้สำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน หมวดคำศัพท์ตำรายาโบราณอีสาน สรุปได้ดังนี้

5.1.1.1 ค้นหาอ่านภาษาไทยถิ่นอีสานด้วยคำปริวรรตไม่ได้

5.1.1.2 คำศัพท์ในคลังศัพท์ส่วนมากเป็นหมวดคำศัพท์ทั่วไป คำศัพท์หมวดตำรายาโบราณอีสานในคลังศัพท์แต่ละเล่มมีจำนวนค่อนข้างน้อย (เล่มที่มีเครื่องหมายดอกจันตรี) คำศัพท์บางคำพบอยู่ในคลังศัพท์ภาษาไทยถิ่นอีสานเพียง 1 เล่มเท่านั้น โดยเฉพาะคำศัพท์ที่เป็นชื่อเฉพาะของชื่อโรค อาการ ชื่อสมุนไพร และโดยส่วนมากไม่มี

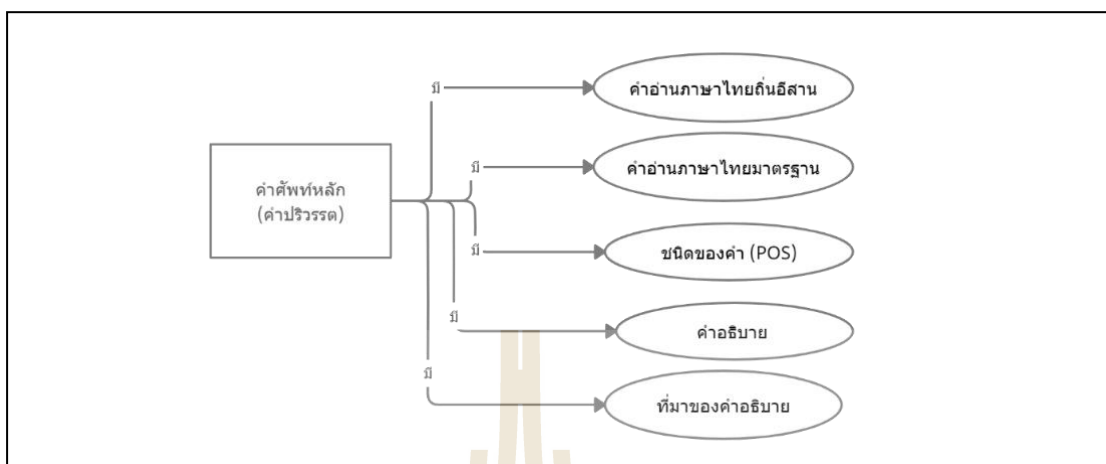
5.1.1.3 การจัดเก็บข้อมูลคำศัพท์อยู่ในรูปแบบหนังสือ ไม่สามารถนำไปประมวลผลด้วยเครื่องคอมพิวเตอร์ได้

5.1.2 การวิเคราะห์และออกแบบโครงสร้างคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

5.1.2.1 ความต้องการของคลังศัพท์สำหรับการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

จากผลการวิเคราะห์ลักษณะโครงสร้างหรือองค์ประกอบในการอธิบายคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสาน พบว่า คลังศัพท์ที่มีอยู่ในปัจจุบันยังไม่สามารถนำไปใช้ในกระบวนการตัดคำและแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานได้อย่างมีประสิทธิภาพ งานวิจัยนี้จึงทำการออกแบบโครงสร้างหรือองค์ประกอบในการอธิบายคำศัพท์ของคลังศัพท์ที่มีลักษณะเฉพาะสำหรับใช้ในกระบวนการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน คือ มีคำปริวรรตสำหรับใช้ตัดคำ และสามารถค้นหาคำอ่านภาษาไทยถิ่นอีสานจากคำปริวรรตได้ ประกอบด้วย 6 องค์ประกอบ ได้แก่ คำศัพท์หลัก (คำปริวรรต) คำอ่านภาษาไทยมาตรฐาน ชนิดของคำ คำอธิบายคำศัพท์ และที่มาของคำอธิบายคำศัพท์ ดังแสดงในรูปแบบที่

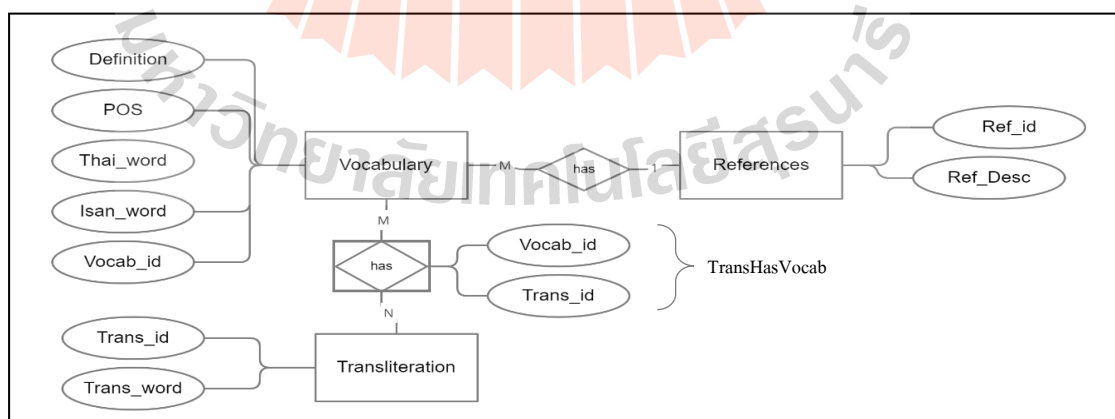
5.1



รูปที่ 5.1 องค์ประกอบการอธิบายคำศัพท์ของคลังศัพท์สำหรับตัดคำและแปลงคำปรีวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

5.1.2.2 การออกแบบโครงสร้างคลังศัพท์ คำปรีวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ประกอบด้วย เอนทิตีและความสัมพันธ์ระหว่างเอนทิตี พจนานุกรมคลังศัพท์ คำปรีวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (Data Dictionary) ชนิดของคำ (Part-Of-Speech: POS) ดังนี้

1) เอนทิตีและความสัมพันธ์ ประกอบด้วยเอนทิตี 4 เอนทิตี ได้แก่ คำศัพท์ (Vocabulary) เอกสารอ้างอิง (Reference) คำปรีวรรต (Transliteration) และเอนทิตี คำปรีวรรต-คำอ่าน (TransHasVocab) รายละเอียดความสัมพันธ์ระหว่างเอนทิตีดังแสดงในรูปที่ 5.2



รูปที่ 5.2 แผนภาพเอนทิตีและความสัมพันธ์ของเอนทิตีในคลังศัพท์ คำปรีวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

2) พจนานุกรมคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน รายละเอียดดังแสดงในตารางที่ 5.2 – 5.5

ตารางที่ 5.2 พจนานุกรมข้อมูลเอ็นทีดีคำศัพท์ (Vocabulary Data Dictionary)

ชื่อแอททริบิวต์	ชนิดของข้อมูล	คำอธิบาย	หมายเหตุ
Vocab_id	Integer	รหัสคำศัพท์	PK, Auto increment
Isan_word	Text	คำอ่านภาษาไทยถิ่นอีสาน	
Thai_word	Text	คำอ่านภาษาไทยมาตรฐาน	
POS	Text	ชนิดของคำ	ใช้ตาม ตารางที่ 5.6
Definition	Text	คำอธิบาย/ความหมายของคำศัพท์	
Ref_id	Integer	รหัสเอกสารอ้างอิง	FK

ตารางที่ 5.3 พจนานุกรมข้อมูลเอ็นทีดีคำปริวรรต (Transliteration Data Dictionary)

ชื่อแอททริบิวต์	ชนิดของข้อมูล	คำอธิบาย	หมายเหตุ
Trans_id	Integer	รหัสคำปริวรรต	PK, Auto increment
Trans_word	Text	คำปริวรรต	

ตารางที่ 5.4 พจนานุกรมข้อมูลเอ็นทีดีแหล่งข้อมูลอ้างอิง (Referenced Resources Data Dictionary)

ชื่อแอททริบิวต์	ชนิดของข้อมูล	คำอธิบาย	หมายเหตุ
Ref_id	Integer	รหัสเอกสารอ้างอิง	PK, Auto increment
Ref_desc	Text	ชื่อเอกสารอ้างอิง	ระบุชื่อเอกสารอ้างอิง

ตารางที่ 5.5 พจนานุกรมข้อมูลเอ็นทีดีคำปริวรรต-คำอ่าน (TransHasVocab Data Dictionary)

ชื่อแอททริบิวต์	ชนิดของข้อมูล	คำอธิบาย	หมายเหตุ
Vocab_id	integer	รหัสคำศัพท์	
Trans_id	integer	รหัสคำปริวรรต	

3) ชนิดของคำ (Part-Of-Speech: POS) การแบ่งชนิดของคำในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน นี้ ใช้กฎทางวากยสัมพันธ์ของภาษาไทยมาตรฐาน (Syntactic Roles) และผลการวิเคราะห์โครงสร้างการเขียนคำราชาโบราณอีสานมาใช้เป็นเกณฑ์ในการกำหนดเป็นชนิดของคำ โดยแบ่งได้เป็น 36 ประเภท ดังแสดงในตารางที่ 5.6

ตารางที่ 5.6 ชนิดของคำในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

No.	POS	คำอธิบาย	ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสาน
1	NPRP	วิสามานยนาม (Proper Noun) หรือ คำนามประเภทชื่อเฉพาะที่ไม่ใช่ชื่อยา ชื่อโรค ชื่อสมุนไพร	ทิพยาร, เทพญาร, ทิพญาร
2	NSMPT	คำนาม หรือ กลุ่มคำ (วลี) ประเภท ชื่อโรค หรือ ชื่อเรียกอาการของโรค	สารบาท, ชะดวง, ไข้มากไม้, ล่องแก้ว, ทกพื้น
3	NPMED	คำนามประเภทชื่อสมุนไพร (พืช, สัตว์, วัตถุ, กระจายยา) และส่วนประกอบ ของพืช หรือสัตว์ ที่นำมาใช้เป็น สมุนไพร	ยานาง, หมอน้อย, บึงเหลื่อม, คันคาก (คางคก), อ้อยดำ, น้ำผึ้ง, น้ำล้าง, ฮากก้วยตีบ, ไบหมากเขียบ, ไม้ไผ่บง
4	NPTBE	คำนามประเภทชนิดของยา (Noun of Pharmacy Type)	น้ำ, ผง, ลูกกลอน, ฝุ่น
5	NTME	คำนามประเภทเวลา (Nouns of Time)	เวลา, เช้า (เช้า), มื้อเช้า (ตอนเช้า), เทียง, เย็น, ค่ำ, แลง, ตะวันชាយ, เฟด, งาม (เช้า), เทียงคืน, กองเด็ก (ตีสี่), ก่อนงาม (ก่อนเที่ยง), ยาม
6	NDAY	คำนามประเภทชื่อวัน (Nouns of Day Name)	วันหนึ่ง, วันสอง, วันสาม, ปะหัส (พฤหัสบดี), ศุกร์, เสาร์, ทิดย์ (อาทิตย์), วันแสม (ข้างแรม), วันเดือนออก (ข้างขึ้น), วัน เดือนดับ (ข้างแรม)
7	NDRT	คำนามประเภทชื่อทิศต่าง ๆ (Noun of Direction)	ตะวันตก, ตะวันออก, เหนือ, ใต้
8	NORG	คำนามประเภทชื่อวัยวะ และ ชื่อองค์ประกอบที่เกิดขึ้นภายในร่างกาย	หัว, ขา, แขน, มือ, เท้า, ท้อง, พื้นดิน, พื้นมือ, หน้าผาก, เลือด, เจริ้วเขี้ยว, เหงื่อ, ใจมือ, ใจ ดิน
9	NMTH	คำนามประเภทชื่อเดือนทางจันทรคติ	เดือน, เดือน๑, อ้าย, ยี่, เจียง, ตี่, ห้า, หก, เจ็ด, แปด, เก้า, สิบ, สิบเอ็ด, สิบสอง,
10	NCNM	คำนามประเภทจำนวนนับ หรือ จำนวน เชิงการนับ (Noun of Cardinal Number)	หนึ่ง, สอง, สาม, ... 1, 2, 3, ... ชาว (ยี่), ชาวห้า (25), ฮ้อย (ร้อย)
11	NORD	คำนามประเภทคำบอกลำดับ	กก (ลำดับแรก), ปลาย (ลำดับสุดท้าย)
12	PDMN	สามัญนาม (Common Noun) หรือ คำนามทั่วไปของ คน พืช สัตว์ และ สิ่งของ	หนังสือยา, พ่อชาย, แม่หญิง, เด็กน้อย, เข้า (ข้าว), ช้าง, ม้า, ไบ, ต้น, หมากไม้, ฮาก, ไบ, ดอก, ไน (เมล็ด), ดอก
13	PPRS	สรรพนามของคน (Personal Pronoun)	เพิ่น, ท่าน, เขา, เฮา, เจ้า, ไผ, สิ่ง

ตารางที่ 5.6 ชนิดของคำในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (ต่อ)

No.	POS	คำอธิบาย	ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสาน
14	PREL	ประพันธสรรพนาม หรือสรรพนามที่ ใช้แทนคำนาม หรือสรรพนามที่กล่าว มาแล้ว (Relative Pronoun)	มัน, อันนั้น
15	VERB	คำกริยาทั่วไป ที่ไม่ใช่กริยาในกลุ่ม อาการป่วย วิธีเตรียมยา วิธีใช้ยา	ไป, ค้าง (วาง), ขุด, ให้, เอา, มี, ฐู่, วาน, เห็น
16	VSTMP	คำกริยาประเภทอาการป่วย (Verb of Symptom)	ปวดหัว, ไข้ (บวม), พวน (มวน), ใอ, ออก ฮ้อน, ฮ้อน, เจ็บ, หาย, สั่น
17	VMTD	คำกริยาประเภทวิธีเตรียมยา (Verb of Method)	ฝน, บด, ตำ, ต้ม, แซ่, คั่ว
18	VDRCT	คำกริยาประเภทวิธีใช้ยา (Verb of Direction)	กิน, อาบ, สม (รม), สูบ, โปะ (โป๊ะ), พุด (พ่น), โสย (โรย)
19	XVRB	กริยาช่วย (Auxiliary Verb)	อย่า, เสีย, ได้, สิ, จะ, ไว้
20	DCNM	คำบอกจำนวน (Determiner, Cardinal Number Expression)	หนึ่ง, อัน๑, บาท๑, ทอกัน (เท่ากัน), นางแล นาง, อันแลอัน (อย่างละอัน), ซุอัน (ทุกอัน), แพ่ง (ริน), ถอง, ทิ้งห้า, แนวแล (อย่างละ)
21	ADV N	คำกริยาวิเศษณ์รูปปกติ (Adverb with Normal Form) ใช้ประกอบหรือขยาย คำกริยาเพื่อให้ได้ใจความชัดเจนและ ละเอียดมากขึ้น	ก่อน, ด้วย, แล้ว
22	ADV I	คำกริยาวิเศษณ์รูปคำซ้ำ (Adverb with Iterative Form)	ก่อน2 (ก่อน ๆ), ทด2 (ทด ๆ)
23	ADJ N	คำคุณศัพท์รูปปกติ (Adjective with Normal Form)	แหม่ม, ไหม้, คำ, แดง, ใหม่, ชุ่น, ใหญ่, น้อย
24	ADJ I	คำคุณศัพท์รูปคำซ้ำ (Adjective with Iterative Form)	แหม่ม2 (แหม่ม ๆ), ไหม้2 (ไหม้ ๆ), แดง2 (แดง ๆ), ขาว2 (ขาว ๆ), ชุ่น ๆ
25	CUNT	ชื่อหน่วยนับ (Unit Classifier)	ก้อน, ต้น, ฝาน, กำ, กลีบ, ดอก, ซีก, บ่วง (ช้อน), บวย
26	CLTV	ลักษณะนามแบบกลุ่ม (Collective Classifier)	ถู่, ชุม (พวก), ฝูง, ทั้งหมด, ทั้งหมด, ทั้งหมด

ตารางที่ 5.6 ชนิดของคำในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (ต่อ)

No.	POS	คำอธิบาย	ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสาน
27	CMTR	ลักษณะนามมาตรวัด (Measurement Classifier)	ตำลึง, บาท, จอมก้อย, เบี้ย, สลึง
28	CFQC	คำบอกลักษณะความถี่ (Frequency Classifier)	ที, เทื่อ (ครั้ง), คาบ (ครั้ง), ครั้ง
29	CCJT	คำสัจฐาน หรือ คำเชื่อม (Coordinating Conjunction)	ผิวา, แล, แลละ, แล้วจึง, เป็นดั่ง, เพราะ, ถ้า, กัน, กันว่า (หากว่า), เท่าว่า (แต่ว่า)
30	PREP	คำบุพบท (Preposition)	ใต้, ล่าง, ที่, เเทิง (บน), แลว, แห่ง, ใน, ตง (ที่), ตาม, ทางใน, ทางลุม, ทางหลัง
31	EAFF	คำจบประโยค (Ending for Affirmative Sentence)	เถอะ, ดีแล, ดูก่อน, เทอญ, แล
32	NEG	คำปฏิเสธ (Negator)	บ่ (ไม่)
33	NEGI	คำปฏิเสธรูปคำซ้ำ (Negator with Iterative Form)	บ่2 (บ่ ๆ)
34	PUNC	เครื่องหมายวรรคตอน (Punctuation)	๑
35	SPELL	คำอา	อัติ วิมัสสมิง, สังกำ พันเต ปาวา
36	UNK	คำที่ยังระบุหมวดคำไม่ได้, คำที่ไม่แน่ใจ (แปลไม่ได้)	ลิลาย

5.1.3 เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

5.1.3.1 การวิเคราะห์เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสาน

ผลการวิเคราะห์ลักษณะของภาษาไทยถิ่นอีสานและเกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสานในคลังศัพท์ภาษาไทยถิ่นอีสาน จำนวน 7 เล่ม (เอกสารชุดเดียวกันกับที่ใช้ในหัวข้อกำหนดปัญหาและความต้องการคลังศัพท์ตำราษาโบราณอีสาน) และตำราษาโบราณอีสาน จำนวน 11 ชุด 459 หน้าลาน พบว่า การสะกดคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสานและตำราษาโบราณอีสานใช้เกณฑ์การสะกดคำ ดังนี้

1) คำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกันแบบพ้องรูปพ้องเสียง หรือคล้ายกับภาษาไทยมาตรฐาน และมีความหมายเดียวกัน การสะกดคำศัพท์ภาษาไทยถิ่นอีสานในตำราษาโบราณอีสานจะสะกดคำศัพท์ตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ในขณะที่การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานจะมีทั้งแบบสะกดคำตรงตามเสียงภาษาไทยถิ่นอีสาน และสะกดคำตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ตัวอย่างดังแสดงในตารางที่ 5.7

ตารางที่ 5.7 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกันแบบพ้องรูปพ้องเสียง หรือคล้ายกับภาษาไทยมาตรฐาน และมีความหมายเดียวกัน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
คน	คน	คน
อาบน้ำ	อาบน้ำ	อาบน้ำ
กิน	กิน	กิน
เป่า	เป่า	เป่า
น้ำอ้อย	น้ำมอ้อย, น้ำอ้อย	น้ำอ้อย
นิ้ว	นิ้ว, นิ้ว	นิ้ว
หน้า	หน้า, หน้า	หน้า
หม้อ	หม้อ, หม้อ	หม้อ
ม่า	ม่า, ม่า	ม่า

2) คำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกัน หรือคล้ายกับภาษาไทยมาตรฐานที่มีตัวสะกดการันต์ และมีความหมายเดียวกัน การสะกดคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาโบราณอีสานจะมีทั้งแบบสะกดคำตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และสะกดคำตรงตามเสียงในภาษาไทยถิ่นอีสาน ไม่มีตัวสะกดการันต์ ตัวอย่างดังแสดงในตารางที่ 5.8

ตารางที่ 5.8 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกัน หรือคล้ายกับภาษาไทยมาตรฐานที่มีตัวสะกดการันต์ และมีความหมายเดียวกัน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
จันทน์แดง	จันแดง, จันทน์แดง	จันแดง*, จันทน์แดง
ประจักษ์	ปะจัก, ประจักษ์	ปะจัก*, ประจักษ์
นมราชสีห์	นมราชสีห์	นมลาชะสี*, นมราชสีห์
จันทน์หอม	จันหอม, จันทน์หอม	จันหอม*, จันทน์หอม
พระจันทร์	พระจันทร์	พะจัน*

หมายเหตุ * สะกดคำตามแบบโบราณ

3) คำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกัน หรือคล้ายกับภาษาไทยมาตรฐานที่เป็นคำควบกล้ำ (ร ล ว) และมีความหมายเดียวกัน การสะกดคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาโบราณอีสานจะมีทั้งสะกดคำตามเสียงภาษาไทยถิ่นอีสาน (ไม่ใช่พยัญชนะควบกล้ำ) และสะกดคำตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ตัวอย่างดังแสดงในตารางที่ 5.9

ตารางที่ 5.9 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงตรงกัน หรือคล้ายกับภาษาไทยมาตรฐานที่เป็นคำควบกล้ำ (ร ล ว) และมีความหมายเดียวกัน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
กว้าง	กวง, กวาง	กวง*, กวาง
ควาย	ควย, ควาย	ควย*, ควาย
พร้อม	พ้อม, พร้อม	พ้อม*, พร้อม
แพร	แพ, แพร	แพ*, แพร
เคือ	เคือ, เคือ	เคือ*, เคือ
ปลาช่อน	ปลาค้อ	ปลาค้อ*
ลูกกลอน	ลูกกอน, ลูกกลอน	ลูกกอน*
กล่าว	ก่าว, กล่าว	ก่าว, กล่าว* (คำต้นฉบับมี ล)
ปล้น	ปลัน	ปลัน (คำต้นฉบับมี ล)
กล้วย	กล้วย	กล้วย*, กล้วย

หมายเหตุ* สะกดคำตามแบบโบราณ

4) คำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงพยัญชนะ “ซ” ซึ่งเป็นปฏิภาค (Correspondence) หรือสอดคล้องกับหน่วยเสียง “ซ” “ซ” “ฌ” หรือ “ส” ในภาษาไทยมาตรฐาน การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานจะมีทั้งแบบสะกดคำตามแบบพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และสะกดคำตามเสียงภาษาไทยถิ่นอีสาน ในขณะที่ตำรายาโบราณอีสานสะกดคำด้วยพยัญชนะ “ซ” “ซ” หรือ “ส” ตัวอย่างดังตารางที่ 5.10

ตารางที่ 5.10 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียงเป็นปฏิภาคกับหน่วยเสียง “ซ” “ช” “ญ” หรือ “ส” ในภาษาไทยมาตรฐาน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
ซ้าง	ซ้าง, ซ้าง	ซ้าง, ซ้าง
ซุ่ม	ซุ่ม, ซุ่ม	ซุ่ม
แซ่	แซ่, แซ่	แซ่, แซ่
ซัง (เกลียด)	ซัง, ซัง	ซัง, ซัง
ชาตรี	ชาตรี, ชาตี	ชาตี*
ฉบับ	ฉบับ, สะฉบับ	ซาฉบับ, สะฉบับ*
เสียบ	เสียบ	เสียบ, เซียบ*
สิ้น	สิ้น, สิ้น	สิ้น
ส่องฟ้า	ส่องฟ้า	ซ่องฟ้า, ส่องฟ้า*
สะอื้น	สะอื้น	ชะอื้น, ซาอื้น*
ชะคราม	ชะคราม, ซ้าคราม	ซ้าคราม, สะคราม*
ชอย	ชอย	สอย
ชื้อ	ชื้อ, ชื้อ	ชื้อ
เช่า	เช่า,เช่า	เช่า

หมายเหตุ * สะกดคำตามแบบโบราณ

5) คำภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “ฮ” ซึ่งเป็นปฏิภาค (Correspondence) หรือความสอดคล้องของเสียงพยัญชนะสอดคล้องกับหน่วยเสียง “ร” ในภาษาไทยมาตรฐาน การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานจะมีทั้งสะกดตรงกับพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (สะกดคำด้วยพยัญชนะ “ร”) และสะกดคำตรงกับเสียงภาษาไทยถิ่นอีสาน (สะกดด้วยพยัญชนะ “ฮ”) ในขณะที่ตำรายาโบราณอีสานสะกดคำตรงกับเสียงภาษาไทยถิ่นอีสาน หรือสะกดคำแบบโบราณ ตัวอย่างดังแสดงในตารางที่ 5.11

ตารางที่ 5.11 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “ฮ” ปฏิภาคกับหน่วยเสียง “ร” ในภาษาไทยมาตรฐาน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
ร้าย	ฮ้าย	ฮ้าย
ราก	ฮาก	ฮาก
ร้อน	ฮ้อน	ฮ้อน
รก	ฮก	ฮก

ตารางที่ 5.11 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “ฮ” ปฏิภาคกับหน่วยเสียง “ร”
ในภาษาไทยมาตรฐาน (ต่อ)

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
เรือน	เฮือน	เฮือน
เรือ่น	เฮือ่น	เฮือ่น
รัง	ฮัง	ฮัง
แรง	แสง	แสง
ดอกกรัก	ดอกกรัก	ดอกฮัก
รุ่งเรือง	สูงเฮือง	สูงเฮือง
รอย	ฮอย	ฮอย

6) คำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง ย-นาสิก (เสียง ย ขึ้นจมูก) การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาโบราณอีสาน ใช้วิธีสะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และสะกดคำตามเสียงภาษาไทยถิ่นอีสาน (ใช้ทั้งพยัญชนะ “ย” และ “ญ”) ตัวอย่างดังแสดงในตารางที่ 5.12

ตารางที่ 5.12 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง ย-นาสิก

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
หญ้า	หญ้ำ	หญ้ำ, หญ้า
หญิง	ญึง	หญึง
ยอด	ยอด	ยอด
ยาง	ยาง	ยาง
ยาก	ยาก	ยาก

7) คำภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “อฮ” การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาโบราณอีสานสะกดคำตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ตัวอย่างดังแสดงในตารางที่ 5.13

ตารางที่ 5.13 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีหน่วยเสียง “อย”

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์ภาษาไทยถิ่นอีสาน (ภาษาไทยถิ่นอีสาน)	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
อย่า	อย่า	อย่า
อยู่	อยู่	อยู่
อย่าง	อย่าง	อย่าง
อยาก	อยาก	อยาก

8) คำศัพท์ภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐานที่มีความหมายเดียวกัน แต่มีหน่วยเสียงสระหรือพยัญชนะแตกต่างกันเล็กน้อย การสะกดคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาโบราณอีสาน มีทั้งสะกดคำตามแบบพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และสะกดคำตามเสียงภาษาไทยถิ่นอีสาน ตัวอย่างดังแสดงในตารางที่ 5.14

ตารางที่ 5.14 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานและภาษาไทยถิ่นอีสานที่มีความหมายเดียวกัน แต่มีหน่วยเสียงสระหรือพยัญชนะต่างกันเล็กน้อย

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์ภาษาไทยถิ่นอีสาน (ภาษาไทยถิ่นอีสาน)	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
เขี้ยว	เงี้ยว, เขี้ยว	เงี้ยว, เขี้ยว
คัดเค้า	คัดเค้า, คัดเค้า	ขัดเค้า, คัดเค้า
ตัวเมียบ	ตัวเมียบ	โตเมียบ, ตัวเมียบ
ลั่นทะเล	ลั่นทะเล	ลั่นทะเล, ลั่นชะเล
สะอึก	สะเออะ	สะเออะ
มะพร้าว	หมากพร้าว	หมากพร้าว, หมากพร้าว
ขมื่น	ขี้หมื่น, ขมื่น	เข้าหมื่น
ข้าวจ้าว	เข้าจ้าว, ข้าวจ้าว	เข้าจ้าว, ข้าวจ้าว

9) คำภาษาไทยถิ่นอีสานที่มีความหมายเดียวกันกับภาษาไทยมาตรฐาน แต่ออกเสียงต่างกัน การสะกดคำในคลังศัพท์ภาษาไทยถิ่นอีสานสะกดคำตามเสียงภาษาไทยถิ่นอีสาน ในขณะที่ตำรายาโบราณอีสานจะมีทั้งแบบรักษาการสะกดคำตามแบบโบราณและสะกดคำตามคลังศัพท์ภาษาไทยถิ่นอีสาน ตัวอย่างดังแสดงในตารางที่ 5.15

ตารางที่ 5.15 ตัวอย่างคำศัพท์ภาษาไทยถิ่นอีสานที่มีความหมายเดียวกันกับภาษาไทยมาตรฐานแต่
ออกเสียงต่างกัน

พจนานุกรมฉบับราชบัณฑิตยสถาน (ภาษาไทยมาตรฐาน)	คลังศัพท์/พจนานุกรม ภาษาไทยถิ่นอีสาน	ตำรายาโบราณอีสาน (ภาษาไทยถิ่นอีสาน)
ลำโพงกาสัก	หมากเขือบ้า	หมากเขือบ้า
ชัน	จู้จู้	จู้จู้
กึ่งไม้	ง่าไม้	ง่าไม้
มะลิ	ดอกซ้อน	ดอกซ้อน
คางคก	คั่นคาก	คั่นคาก
มะเขือพวง	หมากแข้ง	หมากแข้ง, หมากแคว้ง
กำมะถัน	มาด	มาด
จุนสี	สียายอน	สียายอน
หินกรวด	หินแฮ่	หินแฮ่, หินแห่
ดู	สุก	สุก
ขัง	โห่ง	โห่ง
บุกขลิกลึก	อู้จู้	อู้จู้

5.1.3.2 เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังศัพท์ คำปวิวรรต- คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

ผลจากการวิเคราะห์เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสาน สรุปได้
ว่า การสะกดคำอ่านในคลังศัพท์ภาษาไทยถิ่นอีสานและตำรายาอีสานนั้น หากเป็นคำที่มีหน่วยเสียง
ตรงกับภาษาไทยมาตรฐานและเป็นคำที่มีความหมายเดียวกัน ส่วนใหญ่จะใช้เกณฑ์การสะกดคำ
ตรงตามพจนานุกรมฉบับราชบัณฑิต พ.ศ. 2554 ซึ่งจะอ่านออกเสียงไม่ตรงกับภาษาไทยถิ่นอีสาน
และหากเป็นคำภาษาไทยถิ่นอีสาน (ภาษาไทยมาตรฐานออกเสียงอีกแบบหนึ่ง) การสะกดคำมัก
สะกดคำตรงตามเสียงภาษาไทยถิ่นอีสาน การสะกดคำศัพท์จึงมักไม่เป็นมาตรฐานเดียวกัน สำหรับ
งานวิจัยนี้ได้กำหนดเกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังศัพท์ คำปวิวรรต-คำอ่าน
ภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทย โดยใช้วิธีผสมผสานวิธีสะกดคำตามแบบภาษาไทยมาตรฐาน
และการสะกดคำตามเสียงภาษาไทยถิ่นอีสานจังหวัดมหาสารคามและกาฬสินธุ์ (ใช้ตามผู้ปวิวรรต
ตำรายาฉบับที่นำมาศึกษา) มาใช้เป็นเกณฑ์สะกดคำ จำนวน 10 ข้อ รายละเอียดดังแสดงในตาราง
ที่ 5.16

ตารางที่ 5.16 เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังศัพท์ คำปริวรรต-คำอ่าน
ภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสาน	ตัวอย่างการสะกดคำ		
	คำปริวรรต	คำอ่าน ภาษาไทยถิ่นอีสาน	คำอ่าน ภาษาไทยมาตรฐาน
1. คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้อง ความหมายกับคำอ่านภาษาไทยมาตรฐาน ใช้ การสะกดคำตามพจนานุกรมฉบับ ราชบัณฑิตยสถาน พ.ศ. 2554	เห็ด	เห็ด	เห็ด
	เป็น	เป็น	เป็น
	ผักหวาน	ผักหวาน	ผักหวาน
	ดอกไม้	ดอกไม้	ดอกไม้
	ล้าง	ล้าง	ล้าง
2. คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้อง ความหมายกับคำอ่านภาษาไทยมาตรฐาน มี ตัวการ์นต์ ใช้การสะกดคำตามพจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554	ปะจัก	ประจักษ์	ประจักษ์
	จันขาว	จันทน์ขาว	จันทน์ขาว
	จิ้นกอินที	จิ้นกอินทรี	จิ้นกอินทรี
	นมราชาสี	นมราชสีห์	นมราชสีห์
3. คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้อง ความหมายกับคำภาษาไทยมาตรฐาน ที่เป็น คำควบกล้ำ ลว การสะกดคำอ่านภาษาไทย ถิ่นอีสานสะกดคำตามเสียงภาษาไทยถิ่น อีสาน โดยไม่ใช้พยัญชนะควบกล้ำ	กวย	ก้วย	กล้วย
	แพ	แพ	แพร
	ลูกกอน	ลูกกอน	ลูกกลอน
	เกือ	เกือ	เกลือ
	ตะไค	ตะไค้	ตะไคร้
4. คำภาษาไทยถิ่นอีสานมีหน่วยเสียง พยัญชนะ /ส/ ที่เป็นปฏิภาคกับพยัญชนะ / ร/ ในภาษาไทยมาตรฐาน การสะกดคำ อ่านภาษาไทยถิ่นอีสานใช้พยัญชนะ “ส” ตรงกับเสียงในภาษาไทยถิ่นอีสาน	ฮาก	ฮาก	ราก
	เฮือน	เฮือน, เฮือน	เรือน, เรือน
	ฮาง	ฮาง, ฮ้าง	วาง, ร้าง
	ฮอน	ฮอน, ฮ้อน	ร้อน, ร้อน
	ไฮ	ไฮ, ไฮ้	ไร, ไร่
5. คำภาษาไทยถิ่นอีสานมีหน่วยเสียง พยัญชนะ /ซ/ ที่เป็นปฏิภาคกับเสียง /ช/ ในภาษาไทยมาตรฐาน การสะกดคำอ่าน ภาษาไทยถิ่นอีสานใช้พยัญชนะ “ซ” ตรง กับเสียงในภาษาไทยถิ่นอีสาน	ซาง	ซ้าง	ช้าง
	ซื่อ	ซื่อ	ชื้อ
	ซุ่ม	ซุ่ม	ชุ่ม
	เซือก	เซือก	เชือก
6. คำภาษาไทยถิ่นอีสานที่มีคำปริวรรต ประสมด้วยพยัญชนะ /ซ/ /ส/ /ท/ หรือ ประสมด้วยสระที่มีเสียงคล้ายกัน (สระ เสียงสั้น/ยาว) การสะกดคำอ่านภาษาไทย ถิ่นอีสานใช้พยัญชนะและสระตรงกับ เสียงภาษาไทยถิ่นอีสานปัจจุบัน	ชะบับ, ซาบับ	ชะบับ	ฉบับ
	เสียบ, เซียบ	เสียบ	เสียบ
	ลิก, ลีก, ลี๊ก	ลิก, ลี๊ก	ลึก, ลี๊ก
	ชะหนอย, สนอย	ชะหน้อย, สน้อย	สน้อย
	จิ้นกอินที	จิ้นกอินทรี	จิ้นกอินทรี

ตารางที่ 5.16 เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังศัพท์ คำปริวรรต-คำอ่าน
ภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (ต่อ)

เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสาน	ตัวอย่างการสะกดคำ		
	คำปริวรรต	คำอ่าน ภาษาไทยถิ่นอีสาน	คำอ่าน ภาษาไทยมาตรฐาน
7. คำภาษาไทยถิ่นอีสานที่คำปริวรรตประสม คำด้วย พยัญชนะ สระ เสียงคล้ายกันมี ความหมายเดียวกัน การสะกดคำอ่าน ภาษาไทยถิ่นอีสานสะกดคำตามเสียง ภาษาไทยถิ่นอีสานสมัยปัจจุบัน	ผักตำนิน, ผัก ตำนิน	ผักตำนิน	ตำลึง
	เป็ก เป็อก	เป็อก	เปลือก
	จิง, จิง, จิง	จิง, จิง	จิง จริง
	กิ้น กิ้น	กิ้น	กิ้น
	เหล็ง, เหลือง เหล็อง, เหล็ยง	เหล็อง	เหล็อง
	นมลาชาติ, นม ลาตี นมละชะตี	นมราชสีห์	นมราชสีห์
8. คำศัพท์ที่เป็นคำเฉพาะภาษาไทยถิ่นอีสาน ใช้เกณฑ์การสะกดคำตามเสียงภาษาถิ่น อีสานสมัยปัจจุบัน	หมากเขือบา	หมากเขือบ้า	ลำโพงกาซัก
	หินแฮ	หินแห่	หินกรวด
	คันคาก	คันคาก	คางคก
	หอยจวบ	หอยจวบ	หอยขม
	กะทาง	กะท้าง	กิ้งก่ายักษ์
	ผอง	ผอง	ตวง
	สูก	สูก	กู
9. คำภาษาไทยถิ่นอีสานที่คำปริวรรตถ่ายทอด มาจากพยัญชนะต้น “ข หางยาว” อักษรไทย น้อย คำอ่านภาษาไทยถิ่นอีสานใช้รูป พยัญชนะต้นเป็น “อข” ทั้งหมด	อขาก	อขาก	อขาก
	อขู	อขู่	อขู่
	อข่าง	อข่าง	อข่าง
	อข่า	อข่า	อข่า
10. คำภาษาไทยถิ่นอีสานที่มีหน่วยเสียง พยัญชนะต้นเป็น “ข นาสิก” ที่พ้องเสียง พ้องความหมายกับคำภาษาไทยมาตรฐาน ใช้ตัวสะกดตรงตามพจนานุกรมฉบับ ราชบัณฑิตยสถาน พ.ศ. 2554 หากเป็นคำ เฉพาะภาษาไทยถิ่นอีสานให้สะกดคำด้วย “ญ” ทั้งหมด	หยา	หญ้ำ	หญ้ำ
	หยัง	ญิง	หญิง
	ขอม	ญ่อม	ข่อม
	ขาก	ขาก	ขาก
	ขอน	ญ้อน	เพราะ
	หยม	หญุ่ม	ขยม
ยาง	ญา่าง	เดิน	

5.1.4 การพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

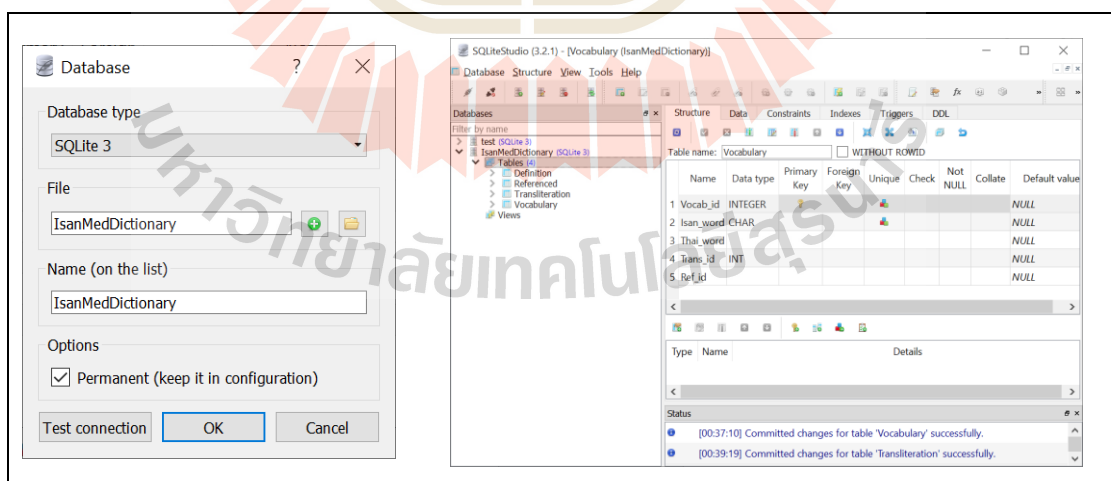
การพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ประกอบด้วย 4 ขั้นตอนหลัก (รูปที่ 5.3) ได้แก่ การสร้างฐานข้อมูลจัดเก็บคำศัพท์ การรวบรวมคำศัพท์ และจัดเตรียมคำศัพท์ การจัดเก็บคำศัพท์ในฐานข้อมูล และประเมินประสิทธิภาพคลังศัพท์



รูปที่ 5.3 ขั้นตอนการพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

ผลการดำเนินการสรุปได้ดังนี้

5.1.4.1 การสร้างฐานข้อมูลคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ได้แก่ การสร้างฐานข้อมูลสำหรับจัดเก็บคำศัพท์ด้วยโปรแกรมระบบจัดการฐานข้อมูล SQLite Studio ตามโครงสร้างคลังศัพท์ที่ได้ออกแบบไว้ในขั้นตอนการวิเคราะห์และออกแบบโครงสร้างคลังศัพท์ ตัวอย่างโครงสร้างฐานข้อมูลคลังศัพท์ดังแสดงในรูปที่ 5.4



รูปที่ 5.4 โครงสร้างฐานข้อมูลคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (IsanMedDictionary) สร้างด้วยระบบจัดการฐานข้อมูล SQLite Studio

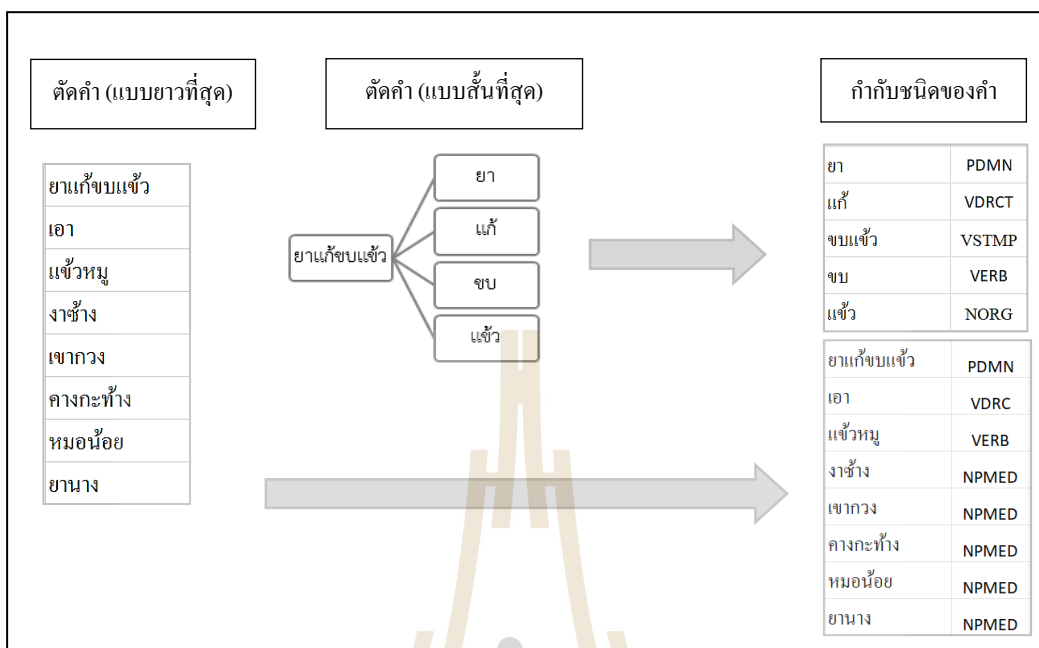
5.1.4.2 การรวบรวม จัดเตรียมคำศัพท์ และจัดเก็บคำศัพท์ในฐานข้อมูล มี 10 ขั้นตอน ดังนี้

1) รวบรวมคำราชาโบราณอีสาน ได้แก่ การนำคำราชาที่ปริวรรตมาจาก คำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ได้แก่ คำราชาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 5 และคำราชาวัดท่าม่วง จังหวัดร้อยเอ็ด เล่ม 2 จำนวนใบลานทั้งสิ้น 11 ฉบับ 459 หน้าลาน ซึ่งบันทึกไว้ในรูปแบบไฟล์เอกสารไมโครซอฟต์เวิร์ด

2) ตัดคำ ได้แก่ การนำคำราชาที่ปริวรรตมาจากคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ได้แก่ คำราชาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 5 และคำราชาวัดท่าม่วง จังหวัดร้อยเอ็ด เล่ม 2 มีจำนวนเอกสารใบลานทั้งสิ้น 11 ฉบับ 459 หน้าลาน ซึ่งบันทึกไว้ในรูปแบบไฟล์เอกสารไมโครซอฟต์เวิร์ดมาตัดคำด้วยมือ โดยรอบแรกใช้เกณฑ์การตัดคำแบบยาวที่สุด และรอบที่ 2 นำเฉพาะคำประสมมาตัดแยกเป็นคำโดด เนื่องจากคำศัพท์ในคำราชาโบราณอีสานมีคำประสมที่สร้างจากคำโดดหลายคำมาประสมกันเป็นคำอยู่เป็นจำนวนมาก เช่น ชื่อโรค ชื่อเรียกอาการของโรค ชื่อสมุนไพร ซึ่งการตัดคำโดยใช้เกณฑ์ตัดคำแบบยาวที่สุดเพียงรูปแบบเดียว ทำให้ได้คำศัพท์ไม่ครอบคลุมคำภาษาไทยถิ่นอีสานที่เป็นคำโดดที่มีอยู่เป็นจำนวนมาก และจะส่งผลกระทบต่อประสิทธิภาพของกระบวนการตัดคำและแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน ดังนั้น งานวิจัยนี้จึงนำคำประสมมาตัดคำเป็นคำโดด เพื่อให้มีจำนวนคำศัพท์ในคลังศัพท์มากขึ้น เช่น “ยาแก้ขบเขว” ตัดคำได้ดังนี้ “ยา” “แก้” “ขบ” “เขว” “ขบเขว” หรือ “ฮิมตากะด้าง” (ขอบตากะด้าง) ตัดคำได้ดังนี้ “ฮิม” “ตา” “กะด้าง” “ฮิมตา” หรือ “บ่กินเข้ากินน้ำ” (ไม่กินข้าวกินน้ำ) ตัดคำได้ดังนี้ “บ่” “กิน” “เข้า” “กิน” “น้ำ” “บ่กินเข้า” เป็นต้น

3) กำกับชนิดของคำ (POS) คือ การนำคำศัพท์ที่ตัดคำแล้วทั้งหมดไปกำกับชนิดของคำ (POS) ในโปรแกรมไมโครซอฟต์เอ็กเซล ตามชนิดของคำในตารางที่ 5.6

ตัวอย่างการตัดคำและกำกับชนิดของคำ ดังแสดงในรูปที่ 5.5



รูปที่ 5.5 ตัวอย่างการตัดคำและกำกับชนิดของคำ

4) ตรวจสอบแก้ตัวสะกดคำอ่านภาษาไทยถิ่นอีสานให้เป็นมาตรฐานเดียวกันตามเกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสานในตารางที่ 5.16 มาเป็นเกณฑ์ในการตรวจและแก้ไขการสะกดคำอ่านภาษาไทยถิ่นอีสาน

5) ทำคำปริวรรต ใช้วิธีแปลงคำอ่านภาษาไทยถิ่นอีสานย้อนกลับเป็นคำปริวรรต โดยใช้เกณฑ์การปริวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบันแบบถ่ายถอดตรงตัวมาใช้เป็นเกณฑ์การแปลงคำอ่านภาษาไทยถิ่นอีสานกลับไปเป็นคำปริวรรต รายละเอียดดังแสดงในตารางที่ 5.17

6) ตรวจสอบคำปริวรรต โดยการเทียบกับภาพตำรายาต้นฉบับ หากไม่ตรงกันจะทำการแก้ไขให้ตรงกัน

ตารางที่ 5.17 เกณฑ์การแปลงคำอ่านภาษาไทยถิ่นอีสานเป็นคำปริวรรต

เกณฑ์การแปลงคำอ่านภาษาไทยถิ่นอีสาน	ตัวอย่าง	
	คำอ่านภาษาไทยถิ่นอีสาน	คำปริวรรต
1. คำอ่านภาษาไทยถิ่นอีสานที่มีวรรณยุกต์คำปริวรรตจะตัดวรรณยุกต์ออกไป	หน่วย	หน่วย
	เข้าจ้าว	เขาจาว
2. คำอ่านภาษาไทยถิ่นอีสานที่มีพยัญชนะควบกล้ำ “ร” “ล” “ว” และตัวการ์นต์ คำปริวรรตจะตัดตัวควบกล้ำและตัวการ์นต์ออกไป	กล้วย	กวย
	จันท์แดง	จันแดง
	กวาง	กวง
	ควายค้อน	ควายคอน
4. คำอ่านภาษาไทยถิ่นอีสานที่ใช้พยัญชนะ “ทร” “ท” “ฉ” คำปริวรรตจะใช้พยัญชนะตรงตามเอกสารต้นฉบับ	จิ้นกอินทรี	จิ้นกอินทรี
	ลิ่นทะเล	ลินทะเล
	ฉบบ	ชะบบ, สะบบ
5. การใช้คำอ่านภาษาไทยถิ่นอีสานที่ประสมด้วยพยัญชนะตัวสะกดแต่ละแม่ คำปริวรรตใช้เกณฑ์ดังนี้		
5.1 แม่ กก (ก ข ค ฉ) คำปริวรรตใช้พยัญชนะ “ก”	มหาเมฆหลวง	มหาเมฆหลวง
5.2 แม่ กค (ค จ ช ซ ฎ ฏ ฐ ฑ ฒ ต ถ ท ธ ศ ษ ส) คำปริวรรตใช้พยัญชนะ “ค”	ทิศ, ประมาท, พิช	ทิศ, ปะมาด, พิด
5.3 แม่ กบ (บ ป ภ พ ฟ) คำปริวรรตใช้พยัญชนะ “บ”	ลิบ	ลิบ
5.4 แม่ กน (น ญ ณ ร ล พ) คำปริวรรตใช้พยัญชนะ “น”	ประมาณ	ปะมาน
5.5 แม่ กง คำปริวรรตใช้พยัญชนะ “ง”	ยาง	ยาง
5.6 แม่ กม คำปริวรรตใช้พยัญชนะ “ม”	หอม	หอม
5.7 แม่ เกย คำปริวรรตใช้พยัญชนะ “ย”	หน่วย	หน่วย
5.8 แม่ เกว คำปริวรรตใช้พยัญชนะ “ว”	ขาว	ขาว
6. คำอ่านภาษาไทยถิ่นอีสานที่ประสมด้วยพยัญชนะต้น “ศ” คำปริวรรตใช้พยัญชนะ “ส”	ศรี	สี
7. คำที่สะกดด้วยอักษรไทยน้อยที่ประสมด้วยสระโอะไม่ม้วน เปลี่ยนเป็นสระโอะไม่มลาย	ให้	ไห
	โก๊, โกล, โก่อ	ไก

7) ลบคำซ้ำ ได้แก่ การใช้คำสั่งลบคำซ้ำของโปรแกรมไมโครซอฟต์เอ็กเซล (Data>Remove Duplicates) ลบคำอ่านภาษาไทยถิ่นอีสานและชนิดของคำที่ซ้ำกันให้เหลือเพียง 1 คำ ตัวอย่างดังแสดงในรูปที่ 5.6

8) กำหนดรายละเอียดคำศัพท์ ได้แก่ การทำคำอ่านภาษาไทยมาตรฐาน คำอธิบาย และแหล่งข้อมูลอ้างอิงที่มาของคำอธิบาย ตัวอย่างดังรูปที่ 5.8

คำปรีวรรค	คำอ่านภาษาอีสาน	คำอ่านไทยมาตรฐาน	POS	คำอธิบาย	อ้างอิง
เขากวาง	เขากวาง	เขากวาง	NP MED	สิ่งที่ยกออกจากรากหัวสัตว์จำพวกกวาง	พจนานุกรมฉบับราชบัณฑิตยฯ
เท่า	เทา	เท่ากัน	DCNM	เสมอกัน	พจนานุกรมฉบับราชบัณฑิตยฯ
เอา	เอา	เอา	VERB	ยึด รับไว้ พา ต้องการ	พจนานุกรมฉบับราชบัณฑิตยฯ
แก้	แก	แก้	VDRCT	ทำให้หาย เช่น แก้โรค แก้เงิน แก้ปัญหา	พจนานุกรมฉบับราชบัณฑิตยฯ
เขี้ยวหมู	เขี้ยวหมู	ฟันหมู	NP MED	ชื่อสมุนไพร	
แล	แล	แล	EAFB	มักใช้ในที่สุกประโยค หรือในที่สุกสร้อยโคลง	พจนานุกรมฉบับราชบัณฑิตยฯ
ขบเขี้ยว	ขบเขี้ยว	กัดฟัน	NSMPT	เอาฟันตอฟันกัดกันไว้แน่น	พจนานุกรมฉบับราชบัณฑิตยฯ
คางกะทัง	คางกะทัง	คางกึ่งก่ายักษ์	NP MED	ดู "กะทัง"	ปรีชา ทิณฑอง
งาช้าง	งาช้าง	งาช้าง	NP MED	1. ชื่อไม้ล้มลุกชนิด <i>Sansevieria cylindrica</i> Bojer ในวงศ์ <i>Dracaenaceae</i> ใบกลมเป็นแท่ง สีเขียวเข้ม ใต้เป็นยาทำให้ อาเจียน, ว่านงาช้าง ก็เรียก 2. ชื่อหอยทะเลหลายชนิดในชั้น <i>Scaphopoda</i> เปลือกลักษณะคล้ายงาช้าง ส่วนมากสีขาว เช่น ชนิด <i>Dentalium aprinum</i> Linn., <i>D. longitrossum</i> Reeve, ฟันช้าง	พจนานุกรมฉบับราชบัณฑิตยฯ

รูปที่ 5.8 ตัวอย่างการกำหนดรายละเอียดคำศัพท์ในคลังศัพท์

ผลการรวบรวมคำศัพท์จากตำรายาโบราณอีสานที่ปรีวรรคมาจากตำรายา อักษรไทยน้อยในใบลานจำนวน 11 ผูก 459 หน้าลาน (ตารางที่ 5.18) ได้คำศัพท์ก่อนการตัดคำซ้ำ ทั้งสิ้น 25,903 คำ และคงเหลือหลังตัดคำซ้ำแล้ว 4,645 คำ หรือ ลดลงคงเหลือร้อยละ 17.93 ของ คำทั้งหมดก่อนตัดคำซ้ำ

ตารางที่ 5.18 สรุปจำนวนคำศัพท์จากตำรายาโบราณอีสานที่ปรีวรรคจากอักษรไทยน้อย

ผูกที่	ตำรายา	จำนวนคำ ทั้งหมด	จำนวนคำ หลังตัดคำซ้ำ	ร้อยละคำศัพท์
1	วัดท่าม่วง เล่ม 2 ฉบับที่ 8	1,416	462	9.95
2	วัดท่าม่วง เล่ม 2 ฉบับที่ 9	2,547	342	7.36
4	วัดท่าม่วง เล่ม 2 ฉบับที่ 10	1,079	320	6.89
3	วัดท่าม่วง เล่ม 2 ฉบับที่ 11	2,649	703	15.14
5	วัดมหาชัย เล่ม 2 ฉบับที่ 1	609	136	2.93
6	วัดมหาชัย เล่ม 2 ฉบับที่ 2	3,789	754	16.24
7	วัดมหาชัย เล่ม 2 ฉบับที่ 3	927	119	2.56
8	วัดมหาชัย เล่ม 2 ฉบับที่ 4	2,318	359	7.73
9	วัดมหาชัย เล่ม 2 ฉบับที่ 5	4,220	618	13.31
10	วัดมหาชัย เล่ม 2 ฉบับที่ 6	4,979	693	14.93
11	วัดมหาชัย เล่ม 2 ฉบับที่ 8	1,370	139	2.99
	รวม	25,903	4,645	17.93

คำศัพท์ที่ได้หลังตัดคำซ้ำจัดกลุ่มตามประเภทความสัมพันธ์ระหว่างคำปริวรรตและคำอ่านภาษาไทยถิ่นอีสาน 4 ประเภท ได้แก่ 1:1 1:M M:1 และ M:N พบว่า ส่วนใหญ่ร้อยละ 61.53 เป็นคำศัพท์ที่มีความสัมพันธ์แบบ 1:1 อันดับสองร้อยละ 36.10 เป็นแบบ M:1 อันดับสามร้อยละ 1.38 เป็นแบบ 1:M และน้อยที่สุดร้อยละ 0.99 คือความสัมพันธ์แบบ M:N (ตารางที่ 5.19)

ตารางที่ 5.19 จำนวนคำศัพท์ภาษาไทยถิ่นอีสานจำแนกตามประเภทความสัมพันธ์

ประเภทความสัมพันธ์	จำนวนคำศัพท์ (คำ)	ร้อยละ
1:1	2,858	61.53
M:1	1,677	36.10
1:M	64	1.38
M:N	46	0.99
รวม	4,645	100

9) ตรวจสอบความถูกต้องของคำศัพท์ในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน โดยให้ผู้เชี่ยวชาญที่มีประสบการณ์การปริวรรตคำราชาโบราณอีสานอักษรไทยน้อยจาก งานอนุรักษ์เอกสารโบราณ สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม ซึ่งเป็นผู้ปริวรรตคำราชาโบราณฉบับที่รวบรวมนำมาสร้างคลังศัพท์ จำนวน 2 คน (ตารางที่ 5.20) เป็นผู้ตรวจสอบความถูกต้อง หลังจากนั้นนำข้อผิดพลาดที่พบมาดำเนินการแก้ไขให้ถูกต้อง

ตารางที่ 5.20 รายชื่อผู้เชี่ยวชาญการปริวรรตคำราชาโบราณอีสานอักษรไทยน้อย

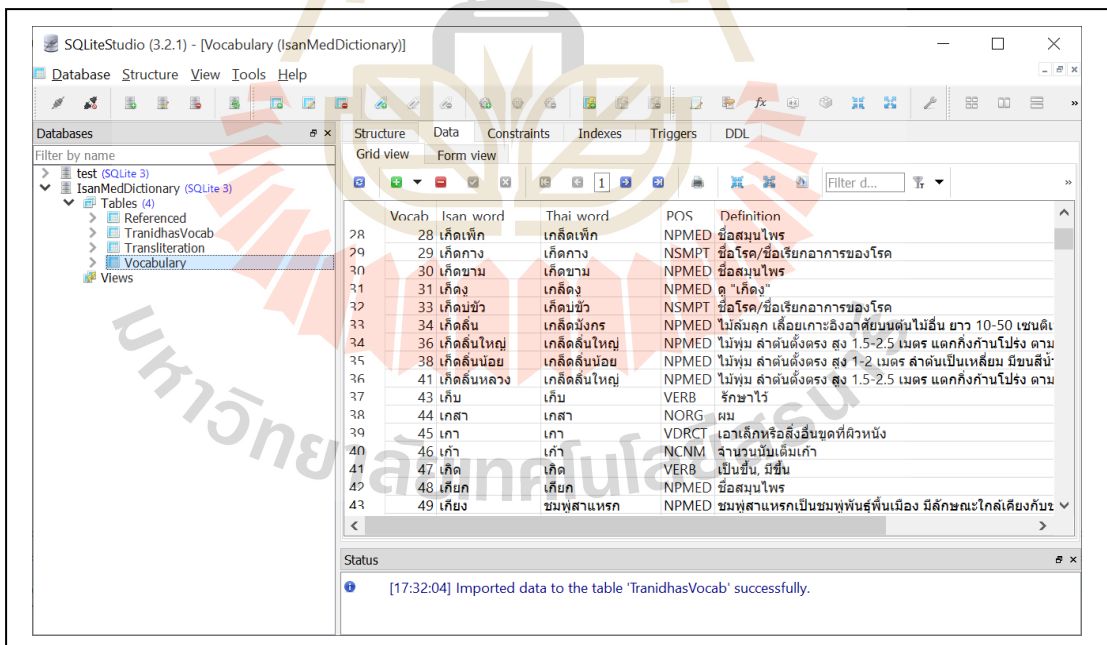
ลำดับ	ชื่อ-สกุล	ตำแหน่ง	ความเชี่ยวชาญด้าน การปริวรรตอักษรโบราณ
1	นายณรงค์ศักดิ์ ราวะรินทร์	นักวิชาการศึกษา	อักษรธรรมอีสาน, อักษรไทยน้อย
2	นายชวนากร จันนาเวช	เจ้าหน้าที่ภาษาโบราณ	อักษรธรรมอีสาน, อักษรไทยน้อย

10) แยกข้อมูลตามโครงสร้างฐานข้อมูล โดยนำข้อมูลมาแบ่งแยกเป็น 4 กลุ่ม ตามโครงสร้างฐานข้อมูลที่ได้ออกแบบเอาไว้ ได้แก่ คำศัพท์ (Vocabulary) เอกสารอ้างอิง (References) คำปริวรรต (Transliteration) และคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน (TransHasVocab) และบันทึกข้อมูลแบบไฟล์ CSV

5.1.4.3 บันทึกจัดเก็บคำศัพท์ลงในฐานข้อมูลคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน โดยการนำรายการคำศัพท์ทั้งหมดที่จัดเก็บอยู่ในรูปแบบเพิ่มข้อมูล CSV แต่ละไฟล์ นำเข้าสู่ฐานข้อมูล ซึ่งสร้างเตรียมเอาไว้แล้วด้วยโปรแกรมระบบจัดการฐานข้อมูล SQLiteStudio ดังรูปที่ 5.9 และตัวอย่างผลลัพธ์ที่ได้หลังจากนำเข้าไฟล์รายการคำศัพท์เข้าสู่ระบบฐานข้อมูลดังแสดงในรูปที่ 5.10



รูปที่ 5.9 ขั้นตอนการนำข้อมูลคำศัพท์เข้าสู่ระบบฐานข้อมูล



รูปที่ 5.10 ตัวอย่างฐานข้อมูลคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

5.1.4.4 การประเมินประสิทธิภาพคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

ประเมิน โดยนักภาษาโบราณ จาก 2 หน่วยงาน คือ กลุ่มหนังสือตัวเขียน และจารึก สำนักหอสมุดแห่งชาติ กรมศิลปากร จำนวน 8 คน และ งานอนุรักษ์เอกสารโบราณ สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม จำนวน 4 คน จำนวนทั้งสิ้น 12 คน รายละเอียดผลการประเมิน (ตารางที่ 5.21) จำแนกเป็น 2 ด้าน ดังนี้

1) ความเหมาะสมของการออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ค่าเฉลี่ยทุกรายการอยู่ในระดับดี $\bar{X} = 3.89$ และ $SD = 0.71$ โดยรายการที่มีค่าเฉลี่ยมากที่สุด คือ ข้อ 1 โครงสร้างคลังศัพท์ที่มีความเหมาะสมกับการนำไปใช้แปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน และคำอ่านภาษาไทยมาตรฐาน อยู่ในระดับดี $\bar{X} = 4.08$ และ $SD = 0.67$ และรายการประเมินที่มีค่าเฉลี่ยน้อยที่สุด คือ ข้อ 3 จำนวนคำศัพท์ มีเพียงพอสำหรับใช้ปริวรรตคำราชาโบราณอีสาน อยู่ในระดับดี $\bar{X} = 3.55$ และ $SD = 0.82$

2) ความเหมาะสมของเกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสาน ค่าเฉลี่ยทุกข้ออยู่ในระดับดี $\bar{X} = 3.91$ และ $SD = 0.72$ โดยเกณฑ์ที่มีความเหมาะสมมากที่สุด คือ ข้อ 4.8 คำปริวรรตที่มีคำอ่านเป็นคำภาษาไทยถิ่นอีสาน สะกดคำตามคลังศัพท์ภาษาไทยถิ่นอีสาน ค่าเฉลี่ยอยู่ในระดับดี $\bar{X} = 4.08$ และ $SD = 0.67$ เกณฑ์ที่มีความเหมาะสมต่ำสุดมี 2 ข้อเท่ากัน คือ ข้อ 4.6 คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานที่ประสมคำด้วยพยัญชนะต้น /ซ/ /ส/ /ท/ และสระที่มีเสียงคล้ายกัน (สระเสียงสั้น/ยาว) การสะกดคำอ่านภาษาไทยถิ่นอีสานใช้พยัญชนะและสระตรงกับภาษาไทยถิ่นอีสานปัจจุบัน และข้อ 4.7 คำปริวรรตที่ประสมคำด้วยพยัญชนะและสระเสียงที่อ่านออกเสียงคล้ายกัน และเป็นคำที่มีความหมายเดียวกัน คำอ่านสะกดคำตามเสียงภาษาไทยถิ่นอีสานปัจจุบัน ค่าเฉลี่ยอยู่ในระดับดี $\bar{X} = 3.75$ และ $SD = 0.62$, , $SD = 0.75$ ตามลำดับ

ตารางที่ 5.21 ผลการประเมินประสิทธิภาพคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

ข้อ	รายการประเมิน	\bar{X}	SD	ประสิทธิภาพ
1	โครงสร้างคลังศัพท์ที่มีความเหมาะสมกับการนำไปใช้แปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน และคำอ่านภาษาไทยมาตรฐาน	4.08	0.67	ดี
2	คำศัพท์ในคลังศัพท์ครอบคลุมคำศัพท์ในหมวดคำราชาโบราณอีสาน	3.83	0.58	ดี
3	จำนวนคำศัพท์มีเพียงพอสำหรับใช้ปริวรรตคำราชาโบราณอีสาน	3.55	0.82	ดี
4	ความเหมาะสมของเกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสาน	3.91	0.72	ดี

ตารางที่ 5.21 ผลการประเมินประสิทธิภาพคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (ต่อ)

ข้อ	รายการประเมิน	\bar{x}	SD	ประสิทธิภาพ
4.1	คำปริวรรตที่มีคำอ่านพ้องเสียงและพ้องความหมายกับคำภาษาไทยมาตรฐาน ใช้การสะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 เช่น “กิน” “ปวด” “ผักหวาน” “ดอกไม้”	3.92	0.51	ดี
4.2	คำปริวรรตที่มีคำอ่านพ้องเสียงและพ้องความหมายกับคำภาษาไทยมาตรฐานที่มีตัวการ์นต์ คำอ่านภาษาไทยถิ่นอีสาน สะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 เช่น “นมราชสีห์” “ประจักษ์” “จันทน์ขาว” “มนต์”	4.00	0.74	ดี
4.3	คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานพ้องเสียงพ้องความหมายกับคำภาษาไทยมาตรฐานที่เป็นคำควบกล้ำ ร ล ว คำอ่านภาษาไทยถิ่นอีสานสะกดคำตามเสียงภาษาไทยถิ่นอีสาน ไม่ใช่พยัญชนะควบกล้ำ เช่น “ปลา” (ปลา) “กล้วย” (กล้วย) “พิก” (พริก) “กวง” (กวาง) “ควย” (ควาย) “เลือ” (เครือ)	3.92	0.72	ดี
4.4	คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานมีหน่วยเสียงพยัญชนะ /ส/ เป็นปฏิภาคกับเสียงพยัญชนะ /ร/ ในภาษาไทยมาตรฐาน การสะกดคำอ่านภาษาไทยถิ่นอีสานใช้พยัญชนะ “ส” ยกตัวอย่าง เช่น “เฮือน” (เรือน) “ฮ้อน” (ร้อน) “ไฮ่” (ไร่)	3.83	0.60	ดี
4.5	คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานมีหน่วยเสียงพยัญชนะ /ซ/ เป็นปฏิภาคกับเสียงพยัญชนะ /ช/ ในภาษาไทยมาตรฐาน การสะกดคำอ่านภาษาไทยถิ่นอีสานใช้พยัญชนะ “ซ” ยกตัวอย่าง เช่น “ซ้าง” (ช้าง) “เชือก” (เชือก) “ซื่อ” (ชื่อ)	4.00	0.60	ดี
4.6	คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานที่ประสมคำด้วยพยัญชนะต้น /ซ/ /ส/ /ท/ และสระที่มีเสียงคล้ายกัน (สระเสียงสั้น/ยาว) การสะกดคำอ่านภาษาไทยถิ่นอีสานใช้พยัญชนะและสระตรงกับภาษาไทยถิ่นอีสานปัจจุบัน เช่น “ซาบับ/ชะบับ” คำอ่านภาษาไทยถิ่นอีสานคือ “ฉบับ” “เซียบ/เสียบ” คำอ่านภาษาไทยถิ่นอีสานคือ “เสียบ” “ชะหน้อย/สน้อย” คำอ่านภาษาไทยถิ่นอีสานคือ “สน้อย” “ซอย/สอย” คำอ่านภาษาไทยถิ่นอีสานคือ “ซอย” (หัน)	3.75	0.62	ดี

ตารางที่ 5.21 ผลการประเมินประสิทธิภาพคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (ต่อ)

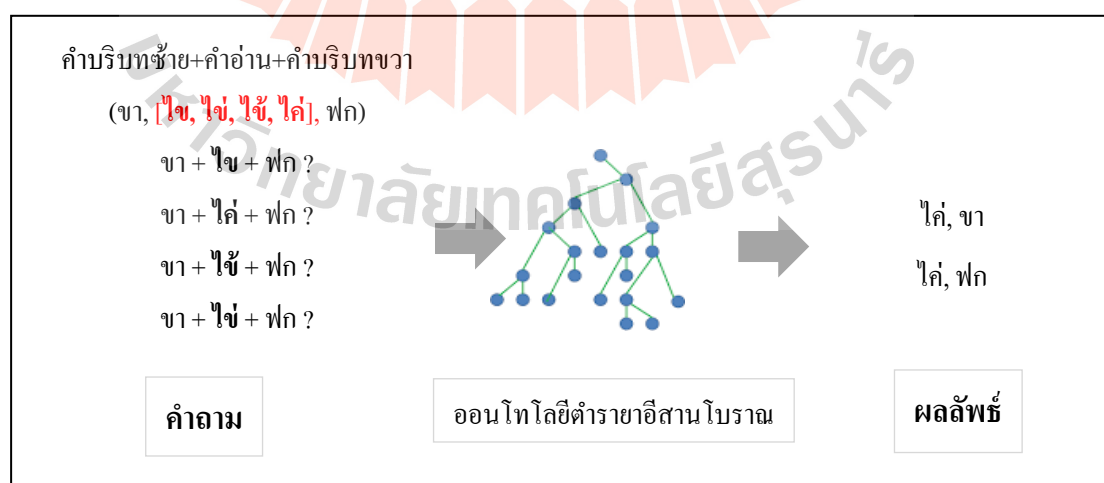
ข้อ	รายการประเมิน	\bar{x}	SD	ประสิทธิภาพ
4.7	คำปริวรรตที่ประสมคำด้วยพยัญชนะและสระเสียงที่อ่านออกเสียงคล้ายกัน และเป็นคำที่มีความหมายเดียวกัน คำอ่านสะกดคำตามเสียงภาษาไทยถิ่นอีสานปัจจุบัน เช่น “กิน/กิน” คำอ่านภาษาไทยถิ่นอีสานเขียนเป็น “กิน” “จึง/จึง” คำอ่านเขียนเป็น “จึง” (จริง)	3.75	0.75	ดี
4.8	คำปริวรรตที่มีคำอ่านเป็นคำภาษาไทยถิ่นอีสาน สะกดคำตามคลังศัพท์ภาษาไทยถิ่นอีสาน เช่น “คั่นคาก” (คางคก) “กะท้าง” (กิ่งก่ายักษ์) “ซูก” (อุ)	4.08	0.67	ดี
4.9	คำอ่านภาษาไทยถิ่นอีสานที่มีคำปริวรรตถ่ายทอดมาจาก “ย หางยาว” อักษรไทยน้อย ที่มีเสียงพยัญชนะต้นเป็นพยัญชนะควบกล้ำ “อช” คำอ่านภาษาไทยถิ่นอีสานใช้รูปพยัญชนะต้นเป็น “อช” ทั้งหมด เช่น “อช่า” “อชู่” “อช่าง” “อชาก”	4.00	0.95	ดี
4.10	คำภาษาไทยถิ่นอีสานที่มีหน่วยเสียงพยัญชนะต้นเป็น ย-นาสิก และพ้องเสียงพ้องความหมายกับคำภาษาไทยมาตรฐาน คำอ่านภาษาไทยถิ่นอีสานสะกดคำตามแบบพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และหากเป็นคำภาษาไทยถิ่นอีสานให้สะกดคำด้วยพยัญชนะ “ญ” ทั้งหมด เช่น คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้องความหมายกับภาษาไทยมาตรฐาน เช่น “หญ่า” “ยา” และคำภาษาไทยถิ่นอีสานที่มีหน่วยเสียงพยัญชนะต้น ย-นาสิก เช่น “ญอง” (ขย) “ญ้อง” (ชม) “ญ้อน” (เพราะ)	4.00	0.60	ดี
5	คลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน นี้ สามารถนำไปใช้สำหรับการปริวรรตคำราชาโบราณอีสานแบบอัตโนมัติได้	3.92	0.90	ดี
6	คำอ่านภาษาไทยถิ่นอีสานและคำอ่านภาษาไทยมาตรฐาน มีความหมายตรงกัน หรือ สอดคล้องกัน	3.75	0.75	ดี
	ค่าเฉลี่ยทุกรายการ	3.89	0.71	ดี

5.2 การพัฒนาออนโทโลยีคำราชาโบราณอีสาน

5.2.1 ความต้องการออนโทโลยีสำหรับแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน

การแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐานนั้น มีคำปริวรรตส่วนหนึ่งที่มีความสัมพันธ์กับคำอ่านภาษาไทยถิ่นอีสานในแบบคำปริวรรตหนึ่งคำมีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำ (1:M) หรือ คำปริวรรตแต่ละคำมีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำอ่าน และคำอ่านภาษาไทยถิ่นอีสานแต่ละคำมีคำปริวรรตได้หลายคำ (M:N) ซึ่งการสืบค้นคำอ่านภาษาไทยถิ่นอีสานของคำปริวรรต 2 ประเภทนี้ด้วยคลังศัพท์จะได้ผลลัพธ์คำอ่านภาษาไทยถิ่นอีสานมากกว่า 1 คำ การเลือกคำอ่านที่ถูกต้อง หรือเหมาะสมที่สุดจะต้องอาศัยบริบท หรือคำที่อยู่รอบข้างของคำนั้นมาช่วยเจาะจงคำอ่านที่ถูกต้อง งานวิจัยนี้จึงพัฒนาออนโทโลยีคำราชาโบราณอีสานเพื่อใช้สอบถามคำอ่านภาษาไทยถิ่นอีสานที่มีความเหมาะสมหรือสอดคล้องกับบริบท สำหรับกรณีคำปริวรรตและคำอ่านภาษาไทยถิ่นอีสานที่มีความสัมพันธ์ในแบบ 1:M และ M:N โดยเฉพาะ

ตัวอย่างดังรูปที่ 5.11 แสดงผลลัพธ์คำอ่านภาษาไทยถิ่นอีสานที่เป็นไปได้ทั้งหมดของคำปริวรรต “ใจ” ได้แก่ “ใจ” “ใจ๋” “ใจ้” “ใจ๊” “ใจ๋” และคำบริบทด้านซ้ายคือ “ขา” คำบริบทด้านขวาคือ “ฟก” เมื่อนำคำอ่านภาษาไทยถิ่นอีสานที่เป็นไปได้ทั้งหมดและคำบริบทไปสอบถามในออนโทโลยีคำราชาโบราณอีสาน จะได้คำตอบคือ มีคำอ่านและคำบริบทที่สอดคล้องกันจำนวน 2 ชุด คือ “ใจ๋” กับ “ขา” และ “ใจ๊” กับ “ฟก” และทั้ง 2 ชุด มีคำอ่านเหมือนกัน แสดงว่าคำอ่าน “ใจ๊” เหมาะสมที่จะนำมาใช้เป็นคำอ่านได้ หากจำนวนผลการค้นหาที่มีจำนวนเท่ากับ 0 หมายถึง ไม่มีคำอ่านที่เหมาะสมหรือไม่ทราบคำอ่านที่เหมาะสม และหากผลลัพธ์คำอ่านมีจำนวนมากกว่า 0 แต่คำอ่านที่ได้ใช้คำเดียวกัน แสดงว่าเป็นคำกำกวมที่มีบริบทเหมือนกัน ไม่สามารถเจาะจงคำอ่านที่เหมาะสมหรือไม่ทราบว่าคำอ่านที่เหมาะสมคือคำใด



รูปที่ 5.11 ตัวอย่างการสอบถามคำอ่านภาษาไทยถิ่นอีสานและบริบทในออนโทโลยีคำราชาโบราณอีสาน

5.2.2 พัฒนาออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน

ผู้วิจัยได้ปรับใช้ออนโทโลยีด้านยาสมุนไพรที่บันทึกไว้ในเอกสารโบราณ (นิสาชล จานงศรี, 2561) ซึ่งประกอบด้วย 12 คลาสหลัก ได้แก่ ยาสมุนไพร เกษัตริย์ รสยา วิธีเตรียมยา วิธีปรุงยา วิธีใช้ยา ผลข้างเคียง ข้อควรระวัง คาถา โรค คน และคัมภีร์โบราณ ซึ่งในงานวิจัยนี้ได้ปรับปรุงโครงสร้างและความสัมพันธ์ระหว่างคลาสใหม่ เพื่อให้ครอบคลุมแนวคิดในการเขียนตำรายาโบราณอีสาน และใช้สำหรับจัดเก็บคำศัพท์ภาษาไทยถิ่นอีสานและความสัมพันธ์กับคำบริบท รายละเอียดผลการพัฒนามีดังนี้

5.2.2.1 ขอบเขตความรู้ในออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน แบ่งเป็น 2 กลุ่ม ตามการจัดกลุ่มความรู้ตามเนื้อหาที่เขียนในตำรายาโบราณอีสาน ได้แก่ เนื้อหาด้านตำรับยา และเนื้อหาด้านอื่น ๆ ที่ไม่ใช่ตำรับยา ดังที่นำเสนอขอบเขตความรู้ด้วยอนุกรมวิธานคำศัพท์ในตำรายาโบราณอีสานในรูปที่ 4.19

5.2.2.2 การออกแบบและสร้างออนโทโลยีตำรายาโบราณอีสาน แบ่งเป็น 20 คลาส ตามโครงสร้างอนุกรมวิธานคำศัพท์ในตำรายาโบราณอีสาน ดังนี้

- 1) Medicine ชื่อยา แบ่งเป็น 3 คลาสย่อย ได้แก่
 - 1.1) HumanMedicineName ชื่อยารักษามนุษย์ เช่น ยาฟก ยามะเอ็ง (ยามะเร.) ยาฝี่ในไส้ ยาแก้พวนใน (ยาแก้มวนท้อง) ยาชักพิษ เป็นต้น
 - 1.2) AnimalMedicineName ชื่อยารักษาสัตว์ เช่น ยาหนอนกินจิ้งกิ้งกวย (หนอนกินจิ้งกวย) ยาจิ้งกวยตาอ่อน (ยาจิ้งกวยตาอ่อน) เป็นต้น
 - 1.3) MedicineType ชนิดยา เช่น ยากิน ยาทา ยาอาบ ยาต้ม ยาฉิม
- 2) Taste รสยา เช่น หวาน ขม ส้ม (เปรี้ยว) จืด
- 3) Symptoms อาการป่วย/ชื่อโรค คำศัพท์ในหมวดนี้ เช่น ไข้ ฮ้อน (ร้อน) ปวดหัว เจ็บท้อง (ปวดท้อง) ล่องแก้วไฟ สารบาด คางเข้ากับ (คางแข็ง) กะยือ (หืด) เป็นต้น
- 4) FormulationMaterial เกษัตริย์ แบ่งเป็น 4 คลาสย่อย ได้แก่
 - 4.1) PlantMaterial พืชวัตถุ แบ่งเป็น 3 หมวดย่อย ได้แก่
 - 4.1.1) PlantMaterialName ชื่อพืชวัตถุ เช่น เกิดลิ้นหลวง (เกิดลิ้นหลวง) เสี้ยวเงินเลียง เป็นต้น
 - 4.1.2) PlantMaterialComponentName ชื่อส่วนประกอบของพืชวัตถุ เช่น ใบ ต้น เปลือก (เปลือก) ฮาก (ราก) หมากพิกกันชัน (พริกชี้ฟ้า) ฮากผักอืดู (รากต้นแมงลัก) ฮากดอกพุด (รากต้นพุด) จี่ชี (ชัน) เป็นต้น

4.1.3) PlantMaterialClassifier ลักษณะนามพืชวัตถุ เช่น กำ
ฝาน หัว ดอก ใบ เป็นต้น

4.2) AnimalMaterial สัตว์วัตถุ แบ่งเป็น 3 คลาสย่อย ได้แก่

4.2.1) AnimalMaterialName ชื่อสัตว์วัตถุ เช่น งู ควาย (ควาย)
ปลา (ปลา) ไก่ดำ คันคาถ (คางครก) กะท้าง (กิ้งก่า) เป็นต้น

4.2.2) AnimalComponentMaterialName ชื่อส่วนประกอบสัตว์
วัตถุ เช่น ดูก (กระดูก) เขา งา ขน ไก่ บี (น้ำดี) หนัง ขี้ควายค่อน (ขี้ควายเผือก) ปิงูเหลื่อม (คิงูเหลื่อม)
 เป็นต้น

4.2.3) AnimalMaterialClassifier ลักษณะนามสัตว์วัตถุ เช่น โด
(ตัว) คาบ (คราบ) ฮัง (รัง) เป็นต้น

4.3) Mineral ธาตุวัตถุ แบ่งเป็น 2 คลาสย่อย ได้แก่

4.3.1) MineralName ชื่อธาตุวัตถุ เช่น หินแท้ (หินกรวด) มาด
(กำมะถัน) ขาง (เหล็กชนิดหนึ่ง) คำ (ทองคำ) เงิน เป็นต้น

4.3.1) MineralClassifier ลักษณะนามของธาตุวัตถุ เช่น ก้อน
 เป็นต้น

4.4) Excipient กระจายยา เช่น น้ำผึ้ง (น้ำผึ้ง) น้ำ น้ำมวก (น้ำแช่
ข้าว) น้ำฝน เหล้า เป็นต้น

5) Pharmacy วิธีเตรียมยาและปรุงยา แบ่งเป็น 3 คลาสย่อย ได้แก่

5.1) Pharmacy วิธีเตรียมยาและปรุงยา เช่น ฝน ดอม ตำ ตาก หลาม
(เผาใส่กระบอก) เป็นต้น

5.2) MedicalMaterial อุปกรณ์ปรุงยา เช่น ถ้วย แพฆาว (ผ้าขาว)
หม้อ บั้ง (กระบอก) เป็นต้น

5.3) FinishingProductForm รูปแบบผลิตภัณฑ์ เช่น ฝุ่น (ผง) น้ำ
ลูกกลอน เป็นต้น

6) Quantity ปริมาณ แบ่ง 2 คลาสย่อย ได้แก่

6.1) Number จำนวนแบบตัวเลข เช่น 1 2 5 7 500

6.2) TextNumber คำบอกปริมาณ เช่น หน่อยหนึ่ง ส่น้อย (สักหน่อย)

ท่อ ประมาณ เป็นต้น

7) Measurement หน่วยวัดและชั่งตวง เช่น บาท ตำลึง ช่อมือ หยิบ เป็นต้น

8) Treatment วิธีใช้ยา เช่น กิน สม (รม) อาบ ทา โปะะ (โป๊ะ) ผิว (ลูบ)

เป็นต้น

- 9) MedicineFees ค่ารักษา เช่น ค่าย ค่า (ทองคำ) เงิน แพรมน (ผ้าเช็ดหน้า)
เป็นต้น
- 10) Patient ผู้ป่วย แบ่งเป็น 2 คลาสย่อย ได้แก่
- 10.1) HumanPatient มนุษย์ แบ่งเป็น 3 คลาสย่อย ได้แก่
- 10.1.1) HumanName ชื่อหรือสรรพนามของมนุษย์ เช่น เขา เจ้า
ผู้ใช้ เฮา เด็กน้อย เป็นต้น
- 10.1.2) HumanSex คำเรียกเพศมนุษย์ เช่น พ่อชาย (ผู้ชาย)
แม่หญิง (ผู้หญิง) เป็นต้น
- 10.1.3) HumanOrgan: ชื่ออวัยวะมนุษย์ เช่น แข้ง ขา แขน มือ
หัว ท้อง เป็นต้น
- 10.2) AnimalPatient สัตว์ แบ่งเป็น 3 คลาสย่อย ได้แก่
- 10.2.1) AnimalName ชื่อสัตว์ เช่น ช้าง (ช้าง) ม้า จั้ว (วัว) ควย
(ควาย) เป็นต้น
- 10.2.2) AnimalSex คำเรียกเพศสัตว์ เช่น ผู้ (ตัวผู้) แม่ (เมีย)
เป็นต้น
- 10.2.3) AnimalOrganName ชื่ออวัยวะสัตว์ เช่น เขา งา หน้
ก้าง ลูก (กระดูก) เป็นต้น
- 11) Rite พิธีกรรม แบ่งเป็น 3 คลาสย่อย ได้แก่
- 11.1) RiteActivity กิจกรรม
- 11.2) RiteMaterial อุปกรณ์หรือเครื่องประกอบพิธีกรรม เช่น โทง
(กระทง) ส่วย (กรวย) ตาแหลหน้าจัว (ฉลอง) เข้าดำ (ข้าวดำ) เข้าแดง (ข้าวแดง) เป็นต้น
- 11.3) Spell คาถา
- 12) DateTime วันเวลา แบ่งเป็น 3 คลาสย่อย ได้แก่
- 12.1) TimeName คำเรียกเวลา เช่น มื้อแลง (ยามค่ำ) มื้อเช้า (ยามเช้า)
งาย (สาย) เทียง ค่ำ ตะวันชาย (ตะวันขึ้น) เป็นต้น
- 12.2) DayName คำเรียกวัน เช่น วันหนึ่ง (วันจันทร์) วันอังคาร
วันปะหัส (พฤหัสบดี) วันศุกร์ วันเสาร์ วันทิด (วันอาทิตย์) วันฟั้ง (วันข้างขึ้น) วันแฮม (วันข้างแรม)
เป็นต้น
- 12.3) MonthName คำเรียกเดือน เช่น เดือนหนึ่ง เดือนสอง เดือนสาม
เดือนเจียง (เดือนอ้าย) เดือนสิบสอง เป็นต้น

13) Verb คำกริยาอื่น ๆ ซึ่งไม่อยู่ในหมวดคำรับยา เช่น สะหั่ว (สระผม) ปั่นหน้า (หันหน้า) เอา ให้ มา ว่า (พูด) จักเอา (จะเอา) อัก (รัก) บูชา (บูชา) ขุด ฟัน (ตัด) เป็นต้น

14) Animal ชื่อเรียกสัตว์และส่วนประกอบของสัตว์ แบ่งเป็น 3 คลาสย่อย ได้แก่

14.1) AnimalName ชื่อสัตว์ เช่น ช้าง (ช้าง) ม้า จั้ว หมู หมา เป็นต้น

14.2) AnimalOrgan ชื่อส่วนประกอบ/อวัยวะของสัตว์ เช่น หนัง เขา จี๋ บี (น้ำดี) เป็นต้น

14.3) AnimalSex ชื่อเรียกเพศสัตว์ เช่น โดผู้ โดแม่ เป็นต้น

15) Noun: คำนามอื่น ๆ แบ่งเป็น 4 คลาสย่อย ได้แก่

15.1) CommonNoun คำนามทั่วไป เช่น หมอ ยา ภูเขา (ยา) เป็นต้น

15.2) Pronoun สรรพนาม เช่น เขา เจ้า แม่เขา พ่อเขา

15.3) DirectionName ชื่อทิศ เช่น กำเหนือ (ทิศเหนือ) กำใต้ (ทิศใต้)

15.4) ClassifierNoun คำนามหน่วยนับ เทื่อ (ครั้ง) พุด (ส่วน) ผั้นแผ่ง เป็นต้น

16) Adverb คำวิเศษณ์ แบ่งเป็น 2 คลาสย่อย ได้แก่

16.1) PreAdverb คำวิเศษณ์หน้ากริยา เช่น ลี (จะ) บ่ (ไม่) ลี (ก่อนข้าง) เป็นต้น

16.2) PostAdverb คำวิเศษณ์หลังกริยา เช่น ทด ๆ ดี นำ (ด้วย) เป็นต้น

17) Adjective คำคุณศัพท์ แบ่งเป็น 2 คลาสย่อย ได้แก่

17.1) DeterminingAdjectives คำบ่งชี้ เช่น นี้ นั้น ไค เป็นต้น

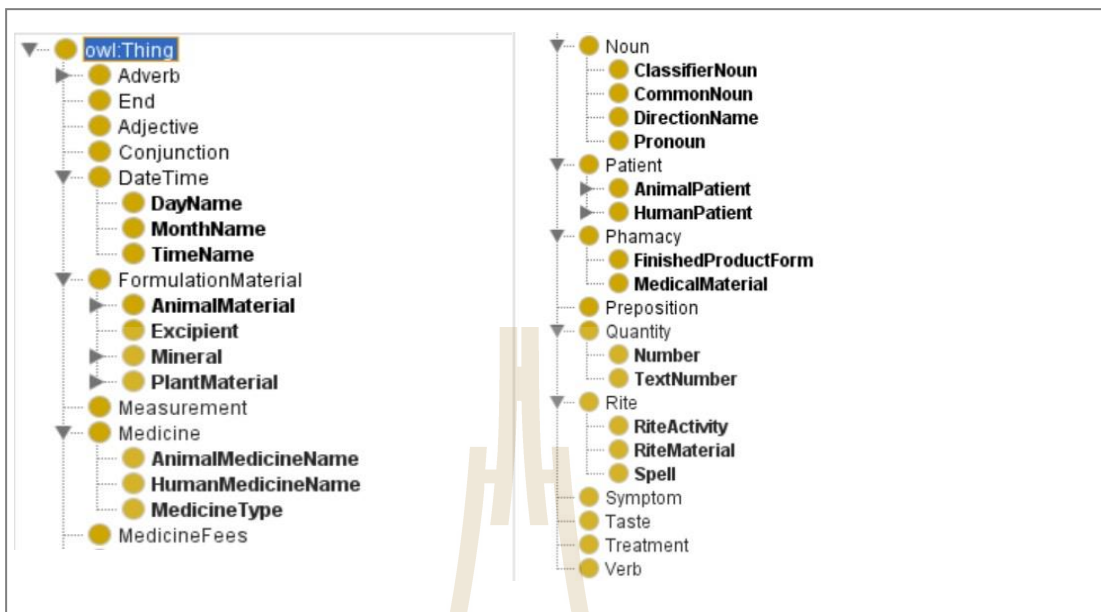
17.2) ClassifierAdjective บอกคุณลักษณะ คำแดง ขาว เหลือง จ้อย (ผอม) เหลือง เป็นต้น

18) Conjunction คำสันธาน เช่น กับ (กับ, และ) คือ ก็ ก็ดี เป็นต้น

19) Preposition คำบุพบท เช่น นำ (ตาม) ทาง (กลาง) ไว้ ข้าง เเทิง (บน, เหนือ) ใต้ (ล่าง) แต่ (จาก) ญื่อน (เพราะ) คั่นว่า (หากว่า) คั่น (หาก) เป็นต้น

20) End คำลงท้ายประโยค เช่น แล ดีแล เทอญ เป็นต้น

ตัวอย่างแผนภาพโครงสร้างลำดับชั้นของคลาสหลักและคลาสย่อยของออนโทโลยีตำรายาโบราณอีสาน ดังแสดงในรูปที่ 5.12



รูปที่ 5.12 โครงสร้างออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน

ตัวอย่างการกำหนดคุณลักษณะของคำศัพท์แต่ละคำ (Properties Assertion) กับคำบริบทด้านซ้ายและขวา ของคำว่า “ใส่” ดังรูปที่ 5.13

รูปที่ 5.13 ตัวอย่างการกำหนดคุณลักษณะของคำศัพท์ (Properties Assertions)

5.2.2.3 Data Properties และ Object Properties ของคลาสต่าง ๆ ในออนโทโลยี
ตำราษาโบราณอีสาน มีรายละเอียดคำอธิบายดังแสดงในตารางที่ 5.22 – 5.23

ตารางที่ 5.22 Data Properties

Data Propertied	คำอธิบาย
PresentIsanWords	คำอ่านภาษาไทยถิ่นอีสาน
StandardThaiWords	คำอ่านภาษาไทยมาตรฐาน
WordId	รหัสคำศัพท์

ตารางที่ 5.23 Object Properties

Object Property Names	คำอธิบาย
has_RightContext	คำศัพท์ที่อยู่ในตำแหน่งติดกันข้างขวา
has_LeftContext	คำศัพท์ที่อยู่ในตำแหน่งติดกันข้างซ้าย

5.2.3 กระบวนการเลือกคำอ่านภาษาไทยถิ่นอีสานที่เหมาะสมด้วยออนโทโลยี

กระบวนการเลือกคำอ่านภาษาไทยถิ่นอีสานที่เหมาะสมด้วยออนโทโลยีคำศัพท์
ตำราษาโบราณอีสาน ประกอบด้วย 2 กระบวนการหลัก ดังนี้

5.2.3.1 นำคำอ่านภาษาไทยถิ่นอีสานที่เป็นผลลัพธ์จากขั้นตอนการค้นหาคำอ่านที่
เป็นไปได้ทั้งหมดในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน
เฉพาะที่มีคำอ่านมากกว่า 1 คำอ่าน และคำบริบทด้านซ้ายและขวาไปสอบถามในออนโทโลยี

5.2.3.2 ขั้นตอนวิธีการเลือกคำอ่านที่เหมาะสม ได้แก่ การนับจำนวนผลลัพธ์
คำอ่านที่ออนโทโลยีแสดงออกมา โดยนับเฉพาะผลลัพธ์คำอ่านที่เหมือนกัน หากจำนวนผลลัพธ์
คำอ่านที่ได้มีจำนวนมากกว่า 1 คำอ่าน (จำนวนคำอ่าน ≥ 1) คำอ่านนั้นจะถูกเลือก นอกเหนือจากนี้
จะใช้คำปริวรรตเป็นคำอ่าน ซึ่งการใช้คำปริวรรตเป็นคำอ่านนี้เป็นเกณฑ์ที่นักภาษาโบราณใช้
สำหรับการทำคำอ่านที่ไม่แน่ใจ หรือไม่ทราบคำอ่านที่ถูกต้อง

ตัวอย่างกระบวนการเลือกคำอ่านภาษาไทยถิ่นอีสานที่เหมาะสมด้วยออนโทโลยี
คำศัพท์ตำราษาโบราณอีสาน ดังแสดงในรูปที่ 5.14

ผลลัพธ์จากขั้นตอน การค้นหาคำอ่านจาก คลังศัพท์	[['ยา', 'ย่า'] 'ปวดท้อง', ['ดิน', 'ดิน'], 'ลง', 'ฮวงลุ่ม', 'แล', 'มดแดง', ['ขา', 'ข่า', 'ข่า'], 'ดำ', ['ไส้', 'ไส้'], 'น้ำอูน', 'ทา']	
คำอ่าน, บริบท (input) ↓	['ยา', 'ย่า'] 'ปวดท้อง' 'ปวดท้อง', ['ดิน', 'ดิน'], 'ลง' 'มดแดง', ['ขา', 'ข่า', 'ข่า'], 'ดำ' 'ดำ', ['ไส้', 'ไส้'], 'น้ำอูน'	
ค้นหาคำอ่านในออน โทโลยีคำศัพท์ (process) ↓	SELECT ?IsanWord ?ThaiWord ?Contex WHERE {{?x med:PresentIsanWords ?IsanWord. FILTER(?IsanWord="ดิน") ?x med:StandardThaiWords ?ThaiWord. ?x med:has_LeftContext ?leftContext. ?leftContext med:PresentIsanWords ?Contex. FILTER(?Contex="ปวดท้อง")}}}	SELECT ?IsanWord ?ThaiWord ?Contex WHERE {{?x med:PresentIsanWords ?IsanWord. FILTER(?IsanWord="ดิน") ?x med:StandardThaiWords ?ThaiWord. ?x med:has_RightContext ?rightContext. ?rightContext med:PresentIsanWords ?Contex. FILTER(?Contex="ลง")}}}
คำอ่าน (output)	ยา, ดิน, ข่า, ไส้	

รูปที่ 5.14 กระบวนการค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยออนโทโลยีคำศัพท์คำยาโบราณอีสาน และตัวอย่าง

5.2.4 ผลการประเมินประสิทธิภาพออนโทโลยีคำศัพท์คำยาโบราณอีสาน

การประเมินประสิทธิภาพของออนโทโลยี ใช้วิธีประเมินจากระดับความถูกต้องของผลลัพธ์คำอ่านภาษาไทยถิ่นอีสานเมื่อใช้ออนโทโลยีเป็นฐานความรู้ในการทำคำอ่าน มีขั้นตอนดังนี้

5.2.4.1 นำประโยคคำรับยาอีสาน โบราณที่คำปริวรรตและคำอ่านภาษาไทยถิ่นอีสานมีความสัมพันธ์กันเป็นแบบ 1:M และ M:N จำนวน 708 ประโยค (1,758 คำ) ที่ได้เตรียมเอาไว้ตามข้อ 3.2.4 การสร้างข้อมูลชุดทดสอบประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน (บทที่ 3 หน้า 74) ไปประมวลผลด้วยโปรแกรมแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานที่เขียนขึ้นตามกรอบการทำงานและขั้นตอนวิธีที่พัฒนาขึ้นมาในงานวิจัยนี้

5.2.4.2 นำผลลัพธ์คำอ่านภาษาไทยถิ่นอีสานที่ได้ไปตรวจความถูกต้องโดยเทียบกับคำอ่านเฉลยที่นักภาษาโบราณทำเอาไว้

5.2.4.3 คำนวณหาค่าร้อยละของความถูกต้อง

ตัวอย่างการตรวจความถูกต้องของผลลัพธ์และเฉลยของผู้เชี่ยวชาญ แสดงในรูปที่ 5.15

ผลลัพธ์ (กรอบการทำงาน)	ยา ปวดท้อง คิน ลง สวงสู่ม แล มดแดง ข่า ตำ ไล่ น้ำอุ่น ทา
เฉลย (ผู้เชี่ยวชาญ)	ยา ปวดท้อง คิน ลง สวงสู่ม แล มดแดง ข่า ตำ ไล่ น้ำอุ่น ทา
ร้อยละความถูกต้อง	$\frac{11}{12} \times 100 = 91.66$

รูปที่ 5.15 ตัวอย่างการตรวจความถูกต้องของคำอ่านและการหาระดับความถูกต้องของคำอ่าน

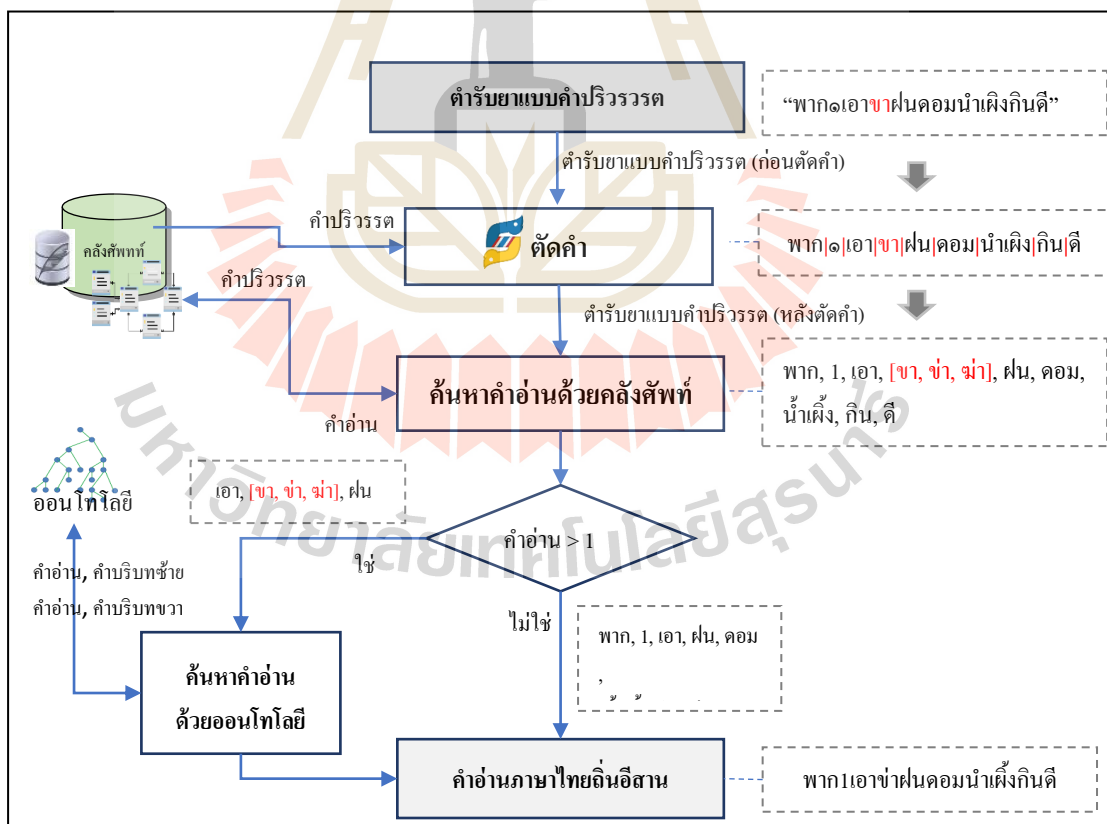
ผลการประเมินประสิทธิภาพการใช้ออนโทโลยีคำศัพท์ตำรายาโบราณอีสานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน พบว่า โดยภาพรวมออนโทโลยีที่พัฒนาขึ้นสามารถค้นคืนคำอ่านได้อย่างถูกต้องตรงกับคำเฉลย ร้อยละ 75.88 โดยประเภทความสัมพันธ์ที่แปลงเป็นคำอ่านที่มีระดับความถูกต้องมากที่สุดคือแบบ M:N จำนวนคำที่นำมาทดสอบทั้งหมด 450 คำ แปลงเป็นคำอ่านถูกต้อง 366 คำ คิดเป็นร้อยละ 81.33 และความสัมพันธ์ที่แปลงเป็นคำอ่านได้ถูกต้องน้อยที่สุดคือแบบ 1:M จำนวนคำที่นำมาทดสอบ 1,308 คำ แปลงเป็นคำอ่านถูกต้อง 968 คำ คิดเป็นร้อยละ 74.01 โดยความผิดพลาดจากกรณีที่เป็นคำที่ไม่รู้จัก (คำที่ไม่มีอยู่ในออนโทโลยี หรือมีคำอ่านในออนโทโลยีแต่คำบริบทไม่ตรงกับออนโทโลยี) คิดเป็นร้อยละ 22.53 และความผิดพลาดกรณีแปลงคำผิด ร้อยละ 1.59 รายละเอียดดังแสดงในตารางที่ 5.24

ตารางที่ 5.24 ผลการประเมินประสิทธิภาพออนโทโลยีตำรายาโบราณอีสาน

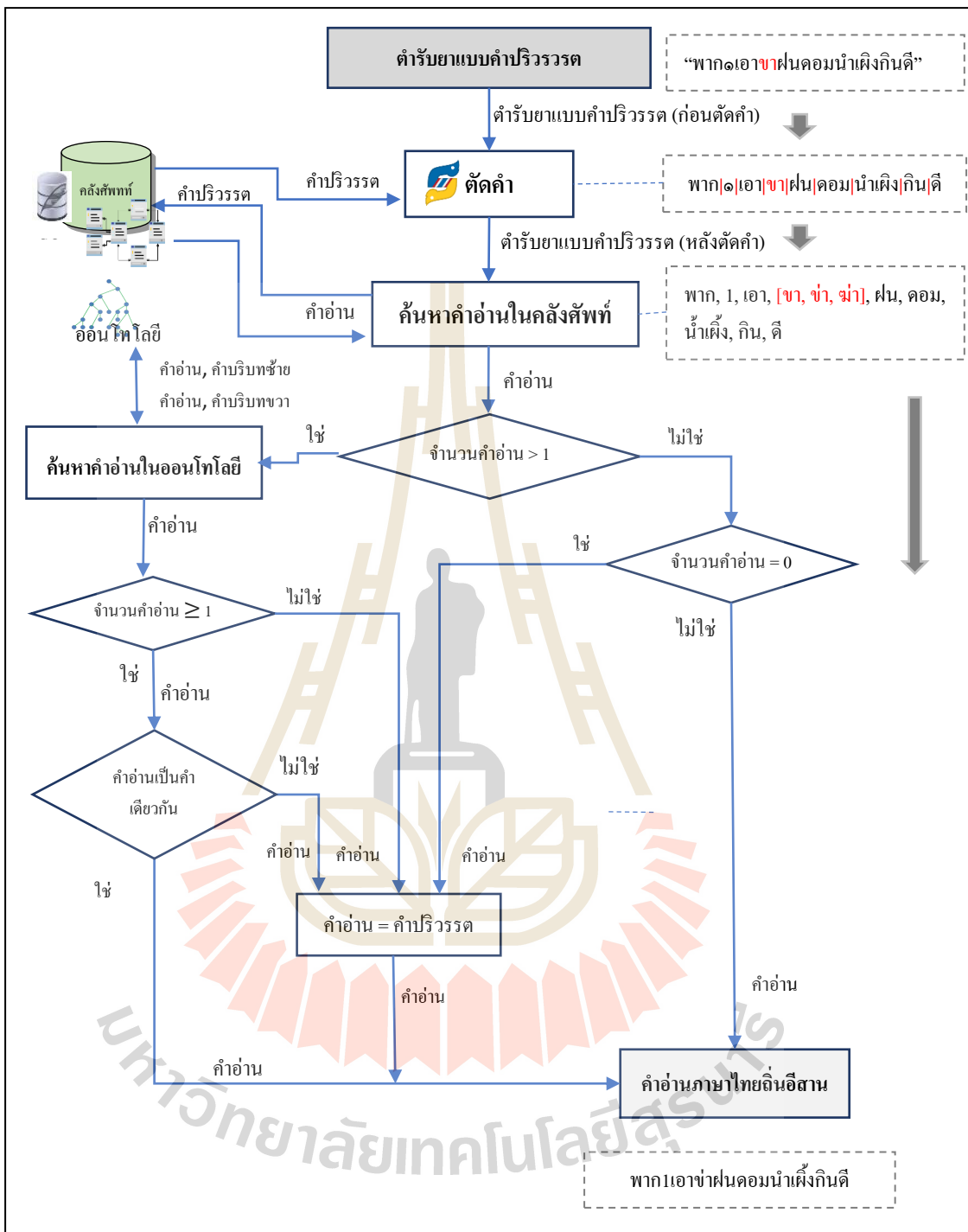
ประเภท	จำนวนคำทดสอบ		คำอ่านถูกต้อง		คำไม่รู้จัก		แปลงคำผิด	
	(คำ)	(ร้อยละ)	คำ	ร้อยละ	คำ	ร้อยละ	คำ	ร้อยละ
1:M	1,308	74.40	968	74.01	319	24.39	21	1.61
M:N	450	25.60	366	81.33	77	17.11	7	1.56
รวม	1,758	100.00	1,334	75.88	396	22.53	28	1.59

5.3 การออกแบบขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

ผลการออกแบบขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน แบ่งเป็น 3 ขั้นตอนหลัก ได้แก่ การตัดคำ การค้นหาคำอ่านด้วยคลังศัพท์ และการค้นหาคำศัพท์ด้วยออนโทโลยี ดังแสดงในรูปที่ 5.16



รูปที่ 5.16 ขั้นตอนหลักของขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน



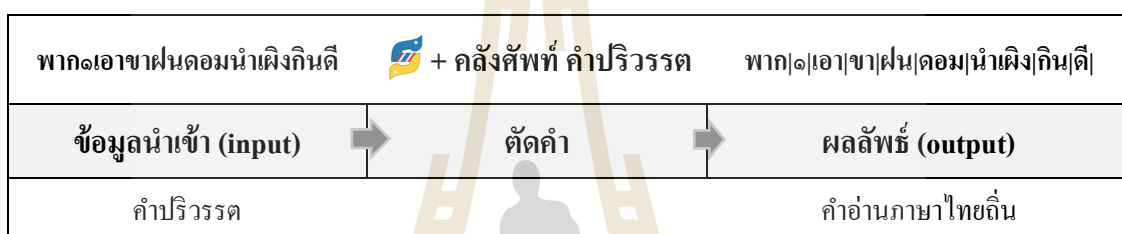
รูปที่ 5.17 ขั้นตอนวิธีการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

รายละเอียดขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน มีดังนี้

5.3.1 การตัดคำ

5.3.1.1 ขั้นตอนการตัดคำ

การตัดคำ คือ การนำเอาประโยคคำรับยาอีสานโบราณแบบคำปริวรรตมาตัดคำด้วยโปรแกรมตัดคำที่เขียนด้วยโปรแกรมภาษาไพธอน (Python) และไลบรารีตัดคำ (Word Tokenizer) ของ PythaiNLP โดยใช้คำปริวรรตจากฐานข้อมูลคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน เป็นคลังศัพท์สำหรับตัดคำ (รูปที่ 5.18)



รูปที่ 5.18 ขั้นตอนการตัดคำปริวรรต

5.3.1.2 ปัญหาในการตัดคำ

การตัดคำโดยใช้รายการคำปริวรรตจากคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน มาตัดคำโดยใช้ไลบรารีตัดคำของ PythaiNLP พบปัญหาในการตัดคำ 2 ประการ ได้แก่

1) การตัดคำผิดพลาด มีสาเหตุจากคำปริวรรตในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ไม่ครอบคลุมคำปริวรรตที่ใช้เขียนตำรายาโบราณอีสานที่เป็นข้อมูลทดสอบ ส่งผลให้คำปริวรรตที่ไม่มีอยู่ในคลังศัพท์เกิดการตัดคำผิดพลาด ดังตัวอย่างรูปที่ 5.19 ซึ่งการตัดคำผิดพลาดจะส่งผลกระทบต่อขั้นตอนการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานจะเกิดความผิดพลาดตามไปด้วย

2) คำปริวรรต 1 คำ ตัดคำได้หลายแบบขึ้นอยู่กับบริบท พบได้ในคำปริวรรตที่เป็นคำประสม หรือคำหลายพยางค์ เช่น “นำทา” ตัดคำเป็นคำประสมคำเดียว คือ “นำทา” (อ่านว่า นำท่า) หรือตัดเป็นคำโดด 2 คำ ได้แก่ “นำ” และ “ทา” (อ่านว่า ทา) หรือ “นำฝน” ตัดเป็นคำโดด 2 คำ ได้แก่ “นำ” (อ่านว่า น้ำ) และ “ฝน” หรือ “นำอุน” (อ่านว่า น้ำอุน) ตัดเป็นคำโดดได้ 2 คำ ได้แก่ “นำ” (อ่านว่า น้ำ) “อุน” (อ่านว่า อุน) เป็นต้น การตัดคำไม่ว่าจะเป็นแบบแยกเป็นคำโดด 2 คำ หรือตัดแบบคำประสมนั้น จะต้องพิจารณาบริบทร่วมด้วยจึงจะทำให้การตัดคำถูกต้อง ตัวอย่างคำปริวรรตที่ตัดคำได้หลายดังรูปที่ 5.20

ก่อนการตัดคำ	หลังการตัดคำ
เจ็บลงต่ำลงแดงลมุกเอาฮากหุงลายหยาโพงโพด ไสบั้งไสบายควนหลามกินดีแลเอาชีกาหลวง ชีกาแดง ทอกันค่านำไว	[เจ็บลงต่ำ, 'ลงแดง', 'ลงมุก', 'เอา', 'ฮากหุงลาย', 'หยา โพงโพด', 'ไสบ', 'บั้ง', 'ไสบ', 'บาย', 'ควน', 'หลาม', 'กิน', 'ดี', 'แล', 'เอา', 'ชีกาหลวง', ' ชีกา ', 'แดง', 'ทอกัน', 'ต่ำ', 'นำ', 'ไว']
ตัดคำผิดพลาด	
สาละบาดไฟ ซอนดังไฟมีกอบนำพวนเอาเปือก เดือหยากผกควยผักหวานบานเขาจาวตำคอมกัน แซกกินดีแลฯ	[สาละ, 'ละ', 'บา', 'ต', 'ไฟ', 'ซอนดังไฟ', 'มัก', 'ออบนำ', 'พวน', 'เอา', 'เปือก', 'เดือ', 'หยากผกควย', 'ผักหวานบาน', 'เขาจาว', 'ตำ', 'คอม', 'กัน', 'แซ', 'กิน', 'ดี', 'แล', 'ฯ']

รูปที่ 5.19 ตัวอย่างการตัดคำปริวรรตผิดพลาด

ผิวขาแหนหอยไหเออฮากหมากมายมาฝนไสน้ำทา ไหกินอันนี้แมนยาแหนหอยแลฯ	ผิวขาตั้งพิคไหเออฮากซาคามาตำไสน้ำทาใจตื่นใจมือ แลผิวนำหัวตัวน้นหายแลฯ
[ผิว, 'ขา', 'แก', 'हनหอย', 'ไห', 'เอา', 'ฮากหมากมาย', 'มา', 'ฝน', 'ไส', 'น้ำทา', 'ไห', 'กิน', 'อันนี้', 'แมน', 'ยา', 'แล', 'हनหอย', 'แล', 'ฯ']	[ผิว, 'ขา', 'ตั้งพิค', 'ไห', 'เอา', 'ฮากซาคามา', 'มา', 'ตำ', 'ไส', 'น้ำ', 'ทา', 'ใจตื่น', 'ใจมือ', 'แล', 'ผิว', 'นำ', 'หัว', 'ตัว', 'น้น', 'หาย', 'แล', 'ฯ']
ตัดคำแบบ 1 คำ 2 พยางค์	ตัดคำแบบคำโดด 2 คำ

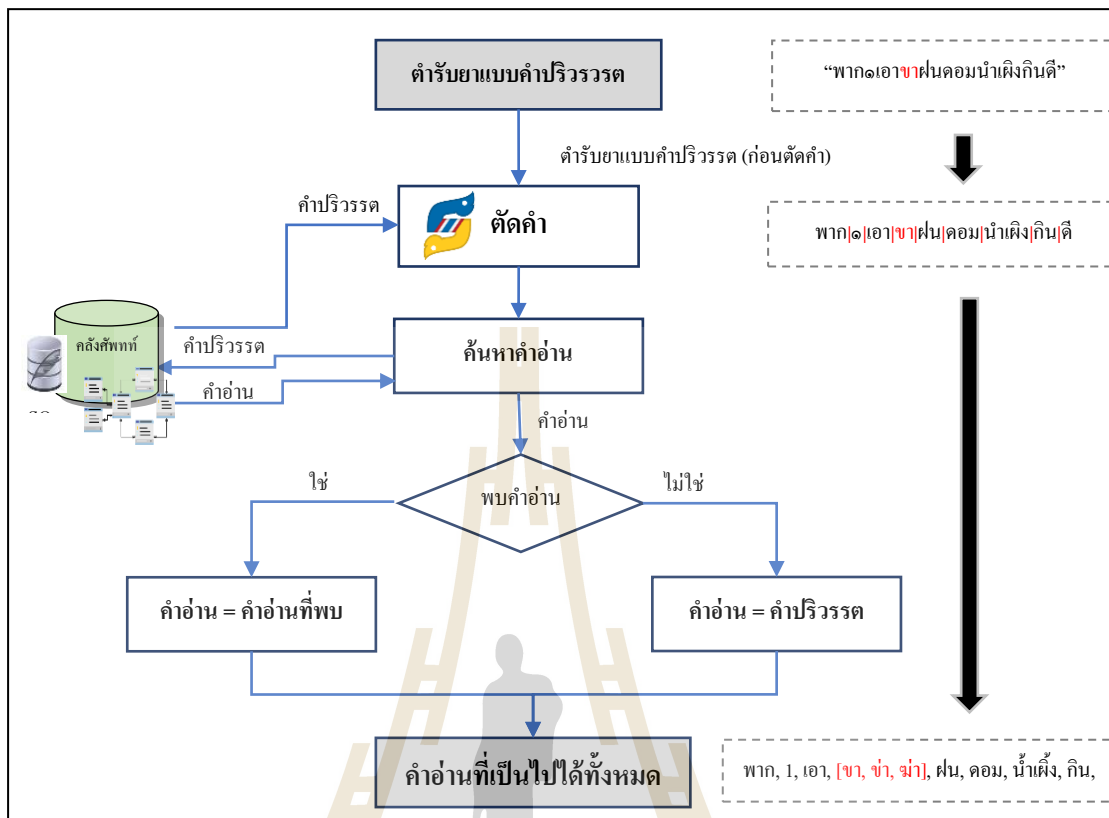
รูปที่ 5.20 ตัวอย่างคำปริวรรตที่สามารถตัดคำได้หลายแบบ

5.3.2 การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์

การค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์ คือ การนำคำปริวรรตที่ตัดคำเรียบร้อยแล้วแต่ละคำไปค้นหาคำอ่านภาษาไทยถิ่นอีสานที่เป็นไปได้ทั้งหมดในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน (รูปที่ 5.21) ซึ่งผลลัพธ์จะมี 2 รูปแบบ ได้แก่

รูปแบบที่ 1 ไม่พบคำอ่านภาษาไทยถิ่นอีสานในคลังศัพท์ (จำนวนคำอ่านภาษาไทยถิ่นอีสาน = 0) กรณีนี้จะใช้คำปริวรรตเป็นคำอ่าน (คำอ่าน = คำปริวรรต)

รูปแบบที่ 2 พบคำอ่านภาษาไทยถิ่นอีสานในคลังศัพท์ (จำนวนคำอ่านภาษาไทยถิ่นอีสาน > 1) คำอ่านทั้งหมดที่ได้นี้จะถูกนำไปสอบถามคำอ่านที่เหมาะสม โดยใช้ออนโทโลยี คำศัพท์คำราชาโบราณอีสานในขั้นตอนถัดไป



รูปที่ 5.21 ขั้นตอนการทำคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์

ตัวอย่างคำรับยาโบราณอีสานแบบคำปริวรรตและผลลัพธ์คำอ่านที่ได้จากกระบวนการค้นหาคำอ่านภาษาไทยถิ่นอีสานโดยใช้คลังศัพท์ ดังรูปที่ 5.22

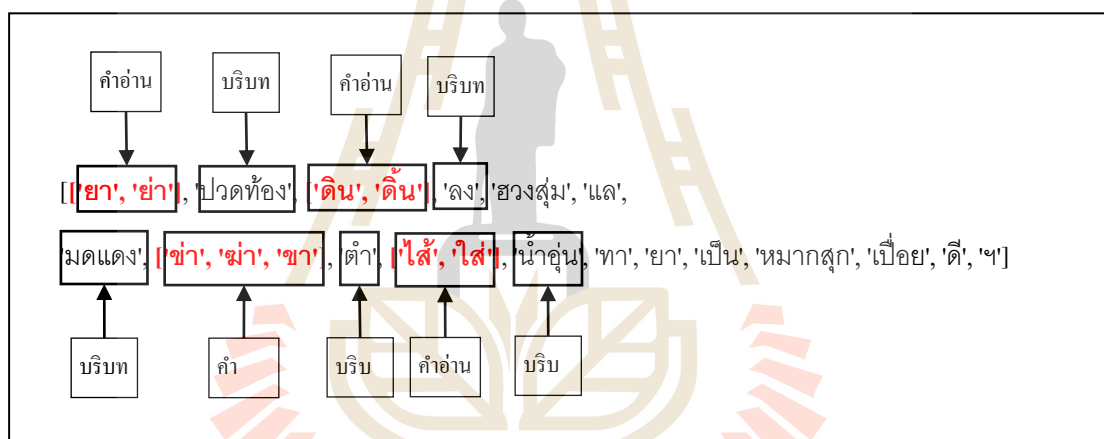
คำปริวรรต	[‘ชา’, ‘ปวดทอง’, ‘ดิน’, ‘ลง’, ‘ฮวงฮุ่ม’, ‘แล’, ‘มดแดง’, ‘ขา’, ‘ตำ’, ‘ไล่’, ‘น้ำอูน’, ‘ทา’, ‘ยา’, ‘เป็น’, ‘หมากสุก’, ‘เป็อย’, ‘ดี’]
คำอ่านภาษาไทยถิ่นอีสาน	[[‘ยา’, ‘ย่า’, ‘ปวดท้อง’, [‘ดิน’, ‘ดิ่น’], ‘ลง’, ‘ฮวงฮุ่ม’, ‘แล’, ‘มดแดง’, [‘ช่า’, ‘ช๋า’, ‘ชา’], ‘ตำ’, [‘ไล่’, ‘ไล่’], ‘น้ำอูน’, ‘ทา’, ‘ยา’, ‘เป็น’, ‘หมากสุก’, ‘เป็อย’, ‘ดี’, ‘๑’]

รูปที่ 5.22 ตัวอย่างผลการค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน

5.3.3 ขั้นตอนวิธีการค้นหาคำอ่านภาษาไทยถิ่นอีสานด้วยออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน

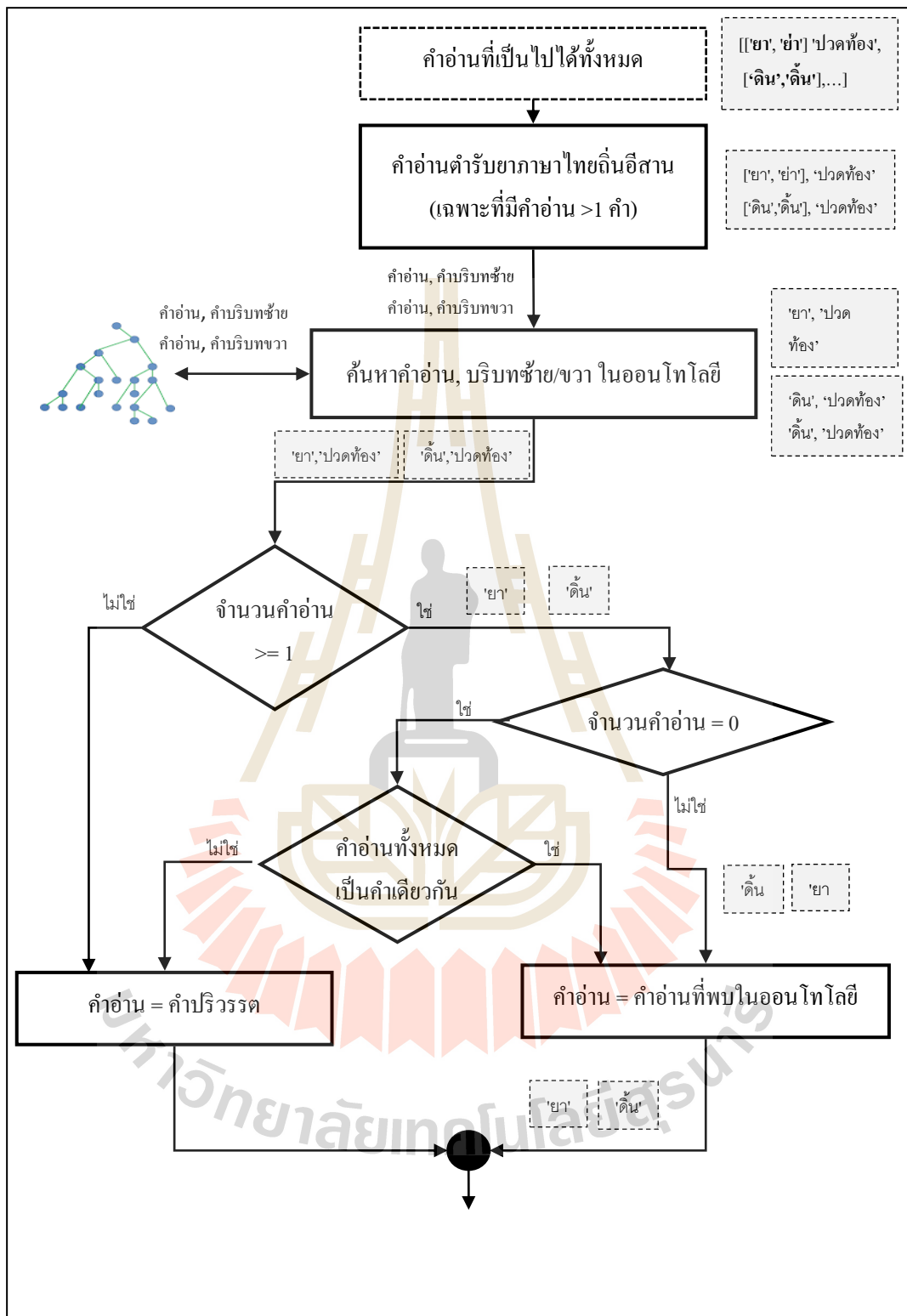
เนื่องจากการค้นหาคำอ่านที่เป็นไปได้ทั้งหมดของคำปริวรรตแต่ละคำด้วยคลังศัพท์ จะมีคำปริวรรตบางคำที่มีคำอ่านได้หลายคำอ่าน ดังนั้น จึงต้องเลือกว่าคำอ่านใดมีความเหมาะสมมากที่สุดที่จะใช้เป็นคำอ่าน ซึ่งงานวิจัยนี้ใช้ออนโทโลยีคำศัพท์ตำรายาโบราณอีสานที่ได้พัฒนาขึ้นมาเพื่อหาคำอ่านที่เหมาะสม โดยการนำคำอ่านแต่ละคำพร้อมคำบริบทด้านซ้ายและขวาที่อยู่ติดกับคำอ่านนั้นข้างละ 1 คำ ไปกำหนดเป็นเงื่อนไขสำหรับการสอบถามในออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน

ตัวอย่างคำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานหลายคำ และคำบริบทซ้าย-ขวาที่จะนำไปสอบถามในออนโทโลยีคำศัพท์ตำรายาโบราณอีสาน ดังรูปที่ 5.23



รูปที่ 5.23 ตัวอย่างคำอ่านภาษาไทยถิ่นอีสานที่มีคำอ่านได้หลายคำ และคำบริบทซ้าย-ขวา

รายละเอียดขั้นตอนวิธีการเลือกคำอ่านที่เหมาะสมด้วยออนโทโลยีคำศัพท์ ดังแสดงในรูปที่ 5.24



รูปที่ 5.24 ขั้นตอนวิธีการเลือกคำอ่านภาษาไทยถิ่นอีสานที่เหมาะสมด้วยออนโทโลยีคำศัพท์

ผลลัพธ์ที่ได้จากการนำเอาคำอ่านและคำบริบทซ้ายและขวาไปสอบถามด้วยออนโทโลยี จะได้ผลลัพธ์ 3 รูปแบบ คือ

รูปแบบที่ 1 จำนวนผลลัพธ์คำอ่านที่ได้จากออนโทโลยี = 1 หมายถึง พบคำอ่านที่เหมาะสม 1 คำ

รูปแบบที่ 2 จำนวนผลลัพธ์คำอ่าน = 0 หมายถึง ไม่พบคำอ่านที่เหมาะสม

รูปแบบที่ 3 จำนวนคำอ่าน ≥ 1 และมีคำอ่านเป็นคำเดียวกันทั้งหมด หมายถึง พบคำอ่านภาษาไทยถิ่นอีสานที่สอดคล้องกับบริบทที่สอบถาม (คำอ่าน = คำอ่านที่พบในออนโทโลยี)

รูปแบบที่ 4 จำนวนคำอ่าน ≥ 1 และมีคำอ่านต่างกัน หมายถึง พบคำอ่านภาษาไทยถิ่นอีสานที่สอดคล้องกับบริบทที่สอบถามหลายคำ จะถือว่าเป็นคำกำกวม ไม่สามารถเจาะจงคำอ่านที่เหมาะสมได้ ในกรณีนี้จะใช้คำปริวรรตเป็นคำอ่าน (คำอ่าน = คำปริวรรต)

5.4 การประเมินประสิทธิภาพกรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

5.4.1 ความถูกต้องแม่นยำ

การประเมินประสิทธิภาพความถูกต้องแม่นยำ (Accuracy) ในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานใช้ข้อมูลชุดทดสอบคือ คำรับยาโบราณอีสานในรูปแบบคำปริวรรต จำนวน 807 คำรับ (7,929 คำ) และเป็นคำรับยาคนละชุดที่นำมาใช้สร้างคลังศัพท์ พบว่า จำนวนคำปริวรรตที่แปลงเป็นคำอ่านภาษาไทยถิ่นอีสานถูกต้องมีจำนวน 6,965 คำ คิดเป็นร้อยละ 87.84 โดยจำแนกตามประเภทความสัมพันธ์ระหว่างคำปริวรรตและคำอ่าน 4 ประเภท เรียงลำดับจากระดับความถูกต้องแม่นยำมากที่สุดถึงน้อยที่สุด ได้แก่ อันดับหนึ่งคือ แบบ 1:1 จำนวนคำที่ทดสอบ 4,396 คำ ถูกต้อง 4,101 คำ (ร้อยละ 93.29) อันดับสอง คือ แบบ M:1 จำนวนคำที่ทดสอบ 1,775 คำ ถูกต้อง 1,530 คำ (ร้อยละ 86.20) อันดับสาม คือ แบบ M:N จำนวนคำที่ทดสอบ 450 คำ ถูกต้อง 366 คำ (ร้อยละ 81.33) และอันดับสี่ คือ แบบ 1:M จำนวนคำที่ทดสอบ 1,308 คำ ถูกต้อง 968 คำ (ร้อยละ 74.01) รายละเอียดดังแสดงในตารางที่ 5.25

ตารางที่ 5.25 ผลการประเมินความถูกต้องของการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

ประเภท	ผลการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน						จำนวนคำทดสอบ	
	ถูกต้อง		คำไม่รู้จัก		แปลงคำผิด			
	คำ	ร้อยละ	คำ	ร้อยละ	คำ	ร้อยละ	คำ	ร้อยละ
1:1	4,101	93.29	292	6.64	3	0.07	4,396	55.44
1:M	968	74.01	319	24.39	21	1.61	1,308	16.50
M:1	1,530	86.20	240	13.52	5	0.28	1,775	22.39
M:N	366	81.33	77	17.11	7	1.56	450	5.67
	6,965	87.84	928	11.80	36	0.45	7,929	100.00

5.4.2 ความผิดพลาด

การแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานด้วยฐานความรู้ประเภทคลังศัพท์และออนโทโลยี จะให้ผลลัพธ์ผิดพลาดใน 2 กรณี ดังนี้

5.4.2.1 คำศัพท์ที่ค้นหาเป็น “คำที่ไม่รู้จัก” หมายถึง คำปริวรรตที่ไม่มีคำอ่านในคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน หรือมีคำอ่านในคลังศัพท์แต่ไม่มีในออนโทโลยีตำราชาโบราณอีสาน (ไม่พบคำอ่านและบริบทตามที่ค้นหาในออนโทโลยี) ซึ่งมีสาเหตุจากคำศัพท์ในคลังศัพท์และออนโทโลยีไม่ครอบคลุมคำศัพท์ที่ค้นหา

5.4.2.2 แปลงคำผิด หมายถึง ผลลัพธ์การแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานไม่ถูกต้อง เกิดจาก 2 สาเหตุ คือ

1) คำปริวรรตที่มีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำ (1:M) แต่ในคลังศัพท์มีเพียง 1 คำอ่าน ทำให้มีตัวเลือกสำหรับนำไปใช้เป็นคำอ่านเพียงตัวเลือกเดียว ตัวอย่างเช่น “หมู” (คำปริวรรต) ควรต้องมีคำอ่านภาษาไทยถิ่นอีสานทั้งหมด 2 คำ ได้แก่ “หมู” และ “หมู่” (พวก) แต่ในคลังศัพท์มีคำอ่านเพียง 1 คำ เท่านั้น คือ “หมู่” ดังนั้น คำปริวรรต “หมู” จึงถูกแปลงเป็นคำอ่านภาษาไทยถิ่นอีสานได้เพียงรูปแบบเดียว คือ “หมู่” เท่านั้น ในขณะที่บางบริบทคำปริวรรตนี้ควรจะต้องแปลงเป็นคำว่า “หมู” จึงจะมีความหมายถูกต้อง ตัวอย่างดังรูปที่ 5.25

ขาม หมู เอาหมากตาเสื่อดินติดหาง หมู บดกิน (คำปริวรรต)
ขาม หมู เอาหมากตาเสื่อดินติดหาง หมู บดกิน (คำอ่านภาษาไทยถิ่นอีสานผิด)
ขาม หมู เอาหมากตาเสื่อดินติดหาง หมู บดกิน (คำอ่านภาษาไทยถิ่นอีสานถูกต้อง)

รูปที่ 5.25 ตัวอย่างการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน กรณี แปลงคำผิด

2) การตัดคำตามเกณฑ์การตัดคำที่ยาวที่สุด ทำให้คำโดดที่เขียนติดกัน ถูกตัดคำแบบคำประสม ซึ่งส่งผลให้ความหมายของคำเปลี่ยนไป เช่น คำปริวรรต “น้ำทา” หากเป็นคำประสม จะมีคำอ่านว่า “น้ำทา” (กระสายยาชนิดหนึ่ง) ขณะที่ในบางบริบทคำนี้ ควรต้องตัดแยกเป็นคำโดด 2 คำ คือ “น้ำ” และ “ทา” (วิธีการใช้ยา) โดยคำว่า “น้ำ” จะมีคำอ่าน 2 แบบ ได้แก่ “น้ำ” (บริเวณ) และ “น้ำ” ดังนั้น การตัดคำที่ยาวที่สุดจึงทำให้เกิดความผิดพลาดในขั้นตอนการทำคำอ่านในบางกรณี ตัวอย่างดังรูปที่ 5.26

บริบทที่ 1 การตัดคำแบบคำประสม	บริบทที่ 2 การตัดคำแบบคำโดด
ยาเมื่อย เอา ก้าน แจ่ว หมุ ชูด บด ใส่ น้ำทา ดี แล ✘	ยาเมื่อย เอา ก้าน แจ่ว หมุ ชูด บด ใส่ น้ำ ทา ดี แล ✔
พาก บ ให้ เอา น้ำทา จัน ทนต์ แดง จูด จี้ ขี้ ตีน คั้ง น้อย ฝน กิน ติ ✔	พาก บ ให้ เอา น้ำ ทา จัน ทนต์ แดง จูด จี้ ขี้ ตีน คั้ง น้อย ฝน กิน ติ ✘
ตัดคำแบบยาวที่สุด	ตัดคำแบบสั้นที่สุด

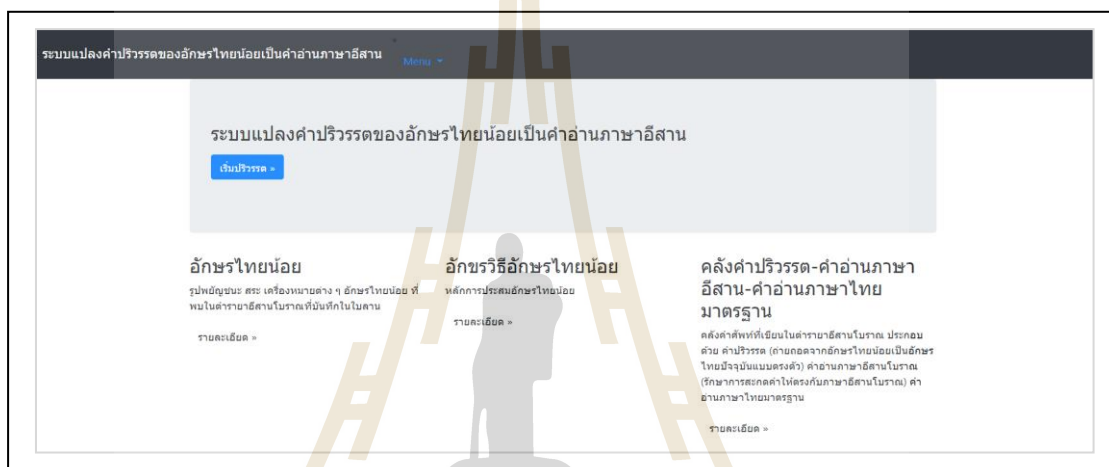
รูปที่ 5.26 ตัวอย่างการตัดคำแบบคำประสมและคำโดดที่ส่งผลต่อคำอ่านและความหมายที่เปลี่ยนไป

5.4.3 ประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

การประเมินประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานนี้ ใช้วิธีประเมินประสิทธิภาพผ่านการใช้ระบบต้นแบบที่พัฒนาขึ้น ขั้นตอนการประเมินประกอบด้วย (1) การพัฒนาระบบต้นแบบ (2) การนำระบบต้นแบบไปให้ให้กลุ่มตัวอย่างทดลองใช้ และ (3) กลุ่มตัวอย่างประเมินประสิทธิภาพ โดยเครื่องมือสำหรับรวบรวมข้อมูลคือ แบบประเมินประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ผลการดำเนินการในแต่ละขั้นตอนมีดังนี้

5.4.2.1 การพัฒนาระบบต้นแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

การวิเคราะห์ ออกแบบ และพัฒนาระบบต้นแบบ ประกอบด้วย 3 ขั้นตอน ได้แก่ (1) วิเคราะห์และออกแบบส่วนติดต่อกับผู้ใช้ตามขั้นตอนการปริวรรตเอกสาร โบราณของนักภาษาโบราณ (2) วิเคราะห์และออกแบบกระบวนการประมวลผลตามกระบวนการและขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานที่ผู้วิจัยพัฒนาขึ้น และ (3) พัฒนาระบบต้นแบบด้วยโปรแกรมภาษาไพธอน (Python) ผลการดำเนินการจะได้หน้าจอหลักดังรูปที่ 5.27



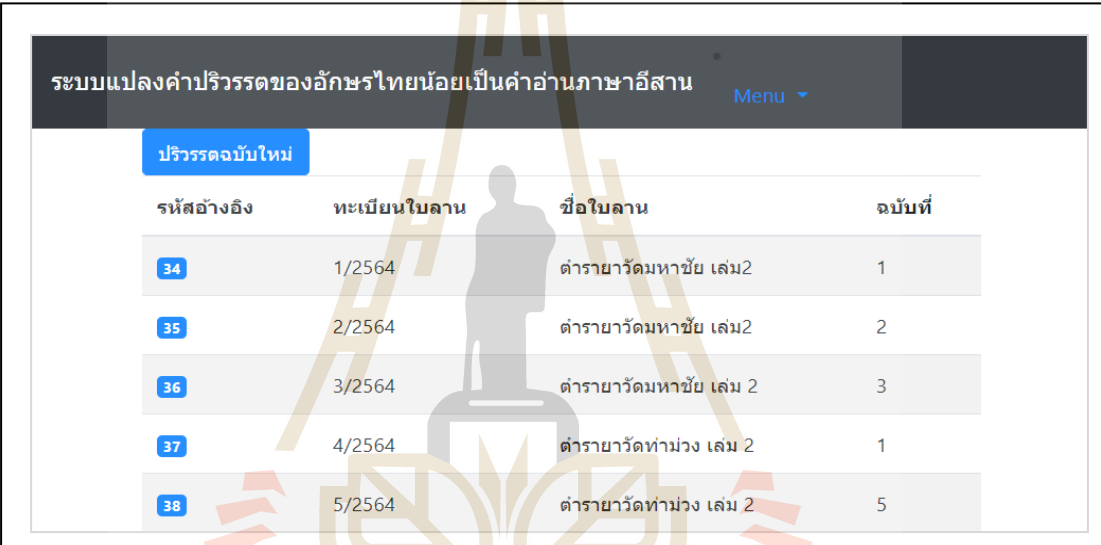
รูปที่ 5.27 ตัวอย่างหน้าจอหลักระบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

จากรูปที่ 5.27 หน้าจอหลักระบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ประกอบด้วยเมนู ดังนี้

- 1) เริ่มปริวรรต คือ เมนูสำหรับการปริวรรตคำรายฉบับใหม่ และจัดการข้อมูลคำรายที่ปริวรรตและจัดทำคำอ่านเสร็จเรียบร้อยแล้ว
- 2) อักษรไทยน้อย คือ เมนูอธิบายลักษณะอักษรไทยน้อยที่ใช้เขียนคำราชาโบราณอีสาน
- 3) อักษรวิธีอักษรไทยน้อย คือ เมนูอธิบายอักษรวิธีการประสมอักษรไทยน้อยเป็นคำในคำราชาโบราณอีสาน
- 4) คลังศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน คือ เมนูค้นหาคำอ่านภาษาไทยถิ่นอีสานและคำอธิบายจากคลังศัพท์

5.4.2.2 การแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ด้วยระบบต้นแบบ แบ่งเป็น 2 ขั้นตอน ได้แก่ การปริวรรตตำรายาโบราณฉบับใหม่ และการจัดการข้อมูลตำรายาที่ได้ทำคำปริวรรตและคำอ่านเสร็จแล้ว มีขั้นตอนการดำเนินการดังนี้

1) การปริวรรตตำรายาฉบับใหม่ เริ่มต้นใช้งานจากหน้าจอหลักโดยคลิกที่เมนู “เริ่มต้นปริวรรต” หลังจากนั้นจะปรากฏหน้าจอที่แสดงเมนู “ปริวรรตฉบับใหม่” และรายชื่อตำรายาที่แปลงคำปริวรรตเป็นคำอ่านเสร็จเรียบร้อยแล้ว ตัวอย่างดังรูปที่ 5.28 แล้วให้คลิกที่เมนู “ปริวรรตฉบับใหม่” เพื่อเรียกฟอร์มกรอกรายละเอียดข้อมูลตำรายาที่จะปริวรรต



รหัสอ้างอิง	ทะเบียนใบลาน	ชื่อใบลาน	ฉบับที่
34	1/2564	ตำรายาวัดมหาชัย เล่ม 2	1
35	2/2564	ตำรายาวัดมหาชัย เล่ม 2	2
36	3/2564	ตำรายาวัดมหาชัย เล่ม 2	3
37	4/2564	ตำรายาวัดท่าม่วง เล่ม 2	1
38	5/2564	ตำรายาวัดท่าม่วง เล่ม 2	5

รูปที่ 5.28 ตัวอย่างหน้าจอปริวรรตตำรายาฉบับใหม่

การกรอกข้อมูลตำรายาฉบับใหม่ มีขั้นตอนดังนี้

1.1) กรอกรายละเอียดข้อมูลของตำรายาฉบับที่จะปริวรรต ได้แก่ หมายเลขทะเบียนใบลาน ชื่อใบลาน (มักตั้งชื่อตามสถานที่พบหรือชื่อเจ้าของใบลาน) จากนั้นกดปุ่ม “บันทึก” เพื่อบันทึกข้อมูลลงในฐานข้อมูล (รูปที่ 5.29)

รูปที่ 5.29 ตัวอย่างหน้าจอสำหรับกรอกรายละเอียดข้อมูลตำรายาที่จะปริวรรตฉบับใหม่

1.2) อัปโหลดไฟล์รูปภาพตำรายาอักษรไทยน้อย (Upload) โดย กดปุ่ม “Choose File” เพื่อเลือกภาพดิจิทัลตำรายาอักษรไทยน้อยที่ต้องการอัปโหลด ประเภทของไฟล์ภาพที่อนุญาตให้อัปโหลดได้แก่ jpeg และ png กรอกรหัสหมายเลขหน้าใบลาน (ลานที่-หน้าที่) และกดปุ่ม “Upload” เพื่ออัปโหลดไฟล์ภาพไปจับเก็บในระบบ (รูปที่ 5.30)

รูปที่ 5.30 ตัวอย่างหน้าจออัปโหลดภาพตำรายาโบราณอีสานที่จะปริวรรต

1.3) ทำคำปริวรรต ได้แก่ การถอดอักษรไทยน้อยเป็นอักษรไทย ปัจจุบันแบบตรงตัว โดยพิมพ์คำปริวรรตลงในช่อง “คำปริวรรต” (รูปที่ 5.31)

1.4) คลิกปุ่ม “ทำคำอ่านภาษาอีสานและภาษาไทยมาตรฐาน” ระบบจะประมวลผลและแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน และภาษาไทยมาตรฐาน แบบอัตโนมัติ โดยจะแสดงคำอ่านในช่องแสดงผล “คำอ่านภาษาไทยถิ่นอีสาน” และ “คำอ่านภาษาไทยมาตรฐาน” ซึ่งหากพบคำอ่านที่ผิดพลาดผู้ปริวรรตสามารถแก้ไขได้ด้วยตนเอง

1.5) กดปุ่ม “Save” เพื่อบันทึกข้อมูล ระบบจะทำการบันทึกไฟล์ภาพดิจิทัลใบลานตำรายาโบราณอีสาน คำปริวรรต คำอ่านภาษาไทยถิ่นอีสาน และคำอ่านภาษาไทยมาตรฐาน ไว้ในฐานข้อมูล

รูปที่ 5.31 ตัวอย่างหน้าจอพิมพ์คำปริวรรต

2) การจัดการข้อมูลตำรายา คือ ขั้นตอนการปรับปรุงข้อมูลตำรายาโบราณอีสานที่แปลงคำปริวรรตเป็นคำอ่านเสร็จเรียบร้อยแล้ว เรียกใช้งานได้จากหน้าจอหลัก โดยคลิกเมนู “เริ่มต้นปริวรรต” ซึ่งจะปรากฏหน้าจอแสดงตำรายาที่เคยปริวรรตและทำคำอ่านเอาไว้แล้ว รายละเอียดประกอบด้วย รหัสอ้างอิง เลขทะเบียนใบลาน ชื่อใบลาน และฉบับที่ (รูปที่ 5.32) แล้วให้คลิกเลือกหมายเลขรหัสอ้างอิงตำรายาฉบับที่ต้องการปรับปรุงแก้ไข

5.4.2.3 การประเมินประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

1) ข้อมูลพื้นฐาน ผู้ประเมิน ประกอบด้วย กลุ่มตัวอย่างที่อ่านอักษรไทยน้อยได้และเข้าใจภาษาไทยถิ่นอีสาน แบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ 1 คือ นักภาษาโบราณในสังกัดกลุ่มหนังสือตัวเขียนและจารึก สำนักหอสมุดแห่งชาติ กรมศิลปากร จำนวน 8 คน และในสังกัด งานอนุรักษ์เอกสารโบราณ สถาบันวิจัยศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม จำนวน 4 คน รวมทั้งสิ้น 12 คน และผู้ประเมินกลุ่มที่ 2 คือ ผู้ใช้ประโยชน์จากตำราภาษาโบราณอีสาน ได้แก่ นักศึกษาด้านสาขาวิชาการแพทย์แผนไทยประยุกต์ ชั้นปีที่ 4 คณะแพทยศาสตรมหาวิทาลัยมหาสารคาม จำนวนทั้งสิ้น 27 คน รายละเอียดดังแสดงในตารางที่ 5.26

ตารางที่ 5.26 ข้อมูลพื้นฐานของผู้ประเมิน

1. กลุ่มตัวอย่าง	นักภาษาโบราณ		ผู้ใช้ประโยชน์จากตำรา
2. หน่วยงาน	กลุ่มหนังสือตัวเขียนและจารึก ฯ หอสมุดแห่งชาติ	งานอนุรักษ์เอกสารโบราณ ฯ มมส.	นักศึกษาด้านวิชาการแพทย์แผนไทยประยุกต์ มมส.
3. ช่วงอายุ	33 - 91 ปี	36 - 50 ปี	22 - 23 ปี
4. ประสบการณ์การปริวรรตเอกสารโบราณ	5 - 57 ปี	6 - 18 ปี	< 1 ปี
5. อักษรโบราณที่เชี่ยวชาญ	- ไทยน้อย - ธรรมอีสาน - ขอม - ธรรมล้านนา - อักษรไทยสมัยสุโขทัย - บัลลวะ	- ธรรมอีสาน - ไทยน้อย - ขอม	- ไทยน้อย - ธรรมอีสาน
6. ภาษาโบราณที่เชี่ยวชาญ	- ภาษาลาว - ภาษาล้านนา - ภาษาไทยโบราณ - บาลี - ลาวหลวงพระบาง	- ภาษาลาว/ ภาษาไทยถิ่นอีสาน - ภาษาไทยโบราณ	

2) ผลการประเมินประสิทธิภาพตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ผู้ประเมินทั้งสองกลุ่มให้ค่าเฉลี่ยทุกรายการอยู่ในระดับดี $\bar{X} = 4.08$ และ $SD = 0.73$ รายการที่มีค่าเฉลี่ยสูงสุดมี 2 รายการเท่ากัน อยู่ในระดับดี ได้แก่ ข้อ 5 ระบบต้นแบบฯ ช่วยให้การสะกดคำอ่านภาษาไทยถิ่นอีสานในตำราอีสานเป็นมาตรฐานเดียวกันมากขึ้น และข้อ 6 ช่วยให้ผู้รู้จักอักษรไทยน้อยแต่ไม่รู้หรือไม่เชี่ยวชาญภาษาไทยถิ่นอีสาน และไม

ผู้เชี่ยวชาญเนื้อหาด้านคำราชาอีสาน สามารถอ่านคำราชาโบราณอีสานอักษรไทยน้อยได้ถูกต้องมากขึ้น $\bar{X} = 4.21$ และ $SD = 0.73$ และ 0.66 ตามลำดับ และรายการประเมินที่มีค่าเฉลี่ยต่ำสุดแต่ยังคงอยู่ในระดับดี คือ ความเร็วในการประมวลผลอยู่ในระดับที่ยอมรับได้ $\bar{X} = 3.64$ และ $SD = 0.84$ เมื่อเปรียบเทียบค่าเฉลี่ยทุกรายการระหว่างกลุ่มตัวอย่าง 2 กลุ่ม พบว่า กลุ่มผู้ใช้ประโยชน์จากคำราชาโบราณอีสานมีค่าเฉลี่ย $\bar{X} = 4.18$ และ $SD = 0.73$ ค่าเฉลี่ยของกลุ่มนักภาษาโบราณ $\bar{X} = 3.85$ และ $SD = 0.67$ ซึ่งกลุ่มผู้ใช้ประโยชน์จากคำราชาโบราณอีสานจะมีค่าเฉลี่ยทุกด้านสูงกว่า รายละเอียดผลการประเมินแต่ละรายการดังแสดงในตารางที่ 5.27

ตารางที่ 5.27 ผลการประเมินประสิทธิภาพขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

ข้อ	รายการประเมิน	นักภาษาโบราณ		ผู้ใช้		รวม 2 กลุ่ม		ประสิทธิภาพ
		\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	
1	ความถูกต้องของผลลัพธ์อยู่ในระดับที่ยอมรับได้	3.83	0.72	4.22	0.70	4.10	0.72	ดี
2	ความเร็วในการประมวลผลอยู่ในระดับที่ยอมรับได้	3.58	0.79	3.67	0.88	3.64	0.84	ดี
3	ระบบต้นแบบฯ ช่วยให้การทำความเข้าใจคำอ่านภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐานทำได้เร็วขึ้น	3.92	0.67	4.22	0.80	4.13	0.77	ดี
4	ระบบต้นแบบฯ ช่วยให้ผู้ทำคำอ่านภาษาไทยถิ่นอีสานและภาษาไทยมาตรฐานมีความถูกต้องมากขึ้น	3.67	0.49	4.26	0.71	4.08	0.70	ดี
5	ระบบต้นแบบฯ ช่วยให้การสะกดคำอ่านภาษาไทยถิ่นอีสานในคำราชาอีสานเป็นมาตรฐานเดียวกันมากขึ้น	3.92	0.67	4.33	0.62	4.21	0.66	ดี
6	ระบบต้นแบบฯ ช่วยให้ผู้รู้จักอักษรไทยน้อยแต่ไม่รู้หรือไม่เชี่ยวชาญภาษาไทยถิ่นอีสาน และไม่เชี่ยวชาญเนื้อหาด้านคำราชาอีสาน สามารถอ่านคำราชาโบราณอีสานอักษรไทยน้อยได้ถูกต้องมากขึ้น	4.08	0.67	4.26	0.76	4.21	0.73	ดี
7	ความพึงพอใจต่อประสิทธิภาพโดยรวมของระบบต้นแบบฯ	3.92	0.67	4.30	0.67	4.18	0.68	ดี
	ค่าเฉลี่ยทุกรายการ	3.85	0.67	4.18	0.73	4.08	0.73	ดี

3) ข้อเสนอแนะ นักภาษาโบราณให้ข้อเสนอแนะว่า การแสดงคำอ่านควรมีการเว้นวรรคเพื่อให้ง่ายต่อการอ่าน คำอ่านภาษาไทยถิ่นอีสานควรรักษาการสะกดคำตามแบบโบราณให้มากที่สุด โดยเฉพาะคำที่เป็นชื่อเฉพาะ ชื่อโรค ชื่ออาการ ชื่อสมุนไพร ควรใช้การสะกดคำอ่านตามแบบโบราณ ไม่ควรเปลี่ยนรูปแบบการสะกดคำตามแบบปัจจุบัน เพื่อให้นำไปใช้ประโยชน์ด้านการศึกษาทางภาษาศาสตร์ได้ด้วย และการแสดงผลคำอ่านควรแสดงแบบบรรทัดต่อบรรทัดตรงกับการเขียนในใบลาน เพื่อให้ง่ายและสะดวกต่อการเปรียบเทียบคำอ่านกับข้อความอักษรไทยน้อยในเอกสารต้นฉบับ คำปวิวรรตควรทำให้สามารถใส่เครื่องหมายสำหรับอธิบายอักษรวิธีการประสมอักษรของคำอักษรโบราณต้นฉบับได้ด้วย สำหรับกลุ่มนักศึกษาไม่มีข้อเสนอแนะ

บทที่ 6

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การใช้ออนโทโลยีเป็นฐานในการแปลงคำปริวรรตอักษรไทยน้อยเป็นคำอ่านภาษาไทยอีสานนี้มีวัตถุประสงค์ (1) เพื่อศึกษาปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย (2) เพื่อออกแบบและพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน และออนโทโลยีคำราชาโบราณอีสาน (3) เพื่อออกแบบและพัฒนาตัวแบบสำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน และ (4) เพื่อประเมินประสิทธิผลและประสิทธิภาพตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน วิธีดำเนินการวิจัยออกเป็น 3 ระยะ ได้แก่ ระยะที่ 1 ศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย ระยะที่ 2 การออกแบบและพัฒนาตัวแบบสำหรับแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน และระยะที่ 3 การประเมินประสิทธิผลของตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน สรุปและอภิปรายผลการวิจัยทั้ง 3 ระยะ มีดังนี้

6.1 สรุปผลการวิจัย

6.1.1 ระยะที่ 1 การศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย

วิธีวิจัย แบ่งเป็น 2 ส่วน ได้แก่ (1) การศึกษากระบวนการปริวรรตเอกสารโบราณ และปัญหาในการปริวรรตคำราชาโบราณอีสานที่บันทึกไว้ในใบลานด้วยอักษรไทยน้อย ใช้วิธีการวิจัยเชิงคุณภาพ เก็บข้อมูลโดยวิธีสัมภาษณ์แบบกึ่งโครงสร้าง กลุ่มตัวอย่าง ได้แก่ นักภาษาโบราณในสังกัดกลุ่มหนังสือตัวเขียนและจารึก หอสมุดแห่งชาติ ท่าวาสุกรี กรุงเทพมหานคร และโครงการอนุรักษ์ใบลานในภาคตะวันออกเฉียงเหนือ มหาวิทยาลัยมหาสารคาม รวมทั้งสิ้น 10 คน วิเคราะห์ข้อมูลโดยการวิเคราะห์ สรุป และจัดกลุ่มประเด็นปัญหา และ (2) การศึกษาโครงสร้างการบันทึกคำราชาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อยที่ส่งผลต่อประสิทธิภาพการปริวรรต ใช้วิธีวิจัยเชิงคุณภาพ กลุ่มตัวอย่าง ได้แก่ คำราชาโบราณอีสานที่ปริวรรตจากคำราชาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อย จำนวน 11 ผูก 459 หน้าลาน ทั้งหมดปริวรรตโดยโครงการอนุรักษ์ใบลานในภาคตะวันออกเฉียงเหนือ มหาวิทยาลัยมหาสารคาม เครื่องมือสำหรับใช้รวบรวมข้อมูล

ได้แก่ แบบวิเคราะห์โครงสร้างการบันทึกคำราชาโบราณอีสาน (3) การศึกษาลักษณะของอักษรไทยน้อย และอักษรวิธีของอักษรไทยน้อยที่ใช้บันทึกคำราชาโบราณอีสานในโบราณ ใชีวิตวิชัยเชิงคุณภาพ กลุ่มตัวอย่าง ได้แก่ สำเนาภาพถ่ายดิจิทัลคำราชาวัดมหาชัย จังหวัดมหาสารคาม เล่ม 1 - 5 และคำราชาวัดท่าม่วง อำเภอเสลภูมิ จังหวัดร้อยเอ็ด เล่ม 1 และเล่ม 2 จากโครงการอนุรักษ์โบราณ ในภาคตะวันออกเฉียงเหนือ มหาวิทยาลัยมหาสารคาม คัดเลือกนำมาศึกษาเฉพาะคำราชาฉบับที่บันทึกด้วยอักษรไทยน้อยที่ทำคำอ่านภาษาไทยถิ่นอีสานเอาไว้แล้ว และสามารถสอบถามข้อมูลจากผู้ปริวรรตได้ จำนวนทั้งสิ้น 11 ผูก 459 ภาพ เครื่องมือสำหรับรวบรวมข้อมูล ได้แก่ แบบวิเคราะห์ลักษณะอักษรไทยน้อยที่ใช้บันทึกคำราชาโบราณอีสานในโบราณ วิเคราะห์ข้อมูลโดยการวิเคราะห์ลักษณะการประสมอักษรเป็นคำด้วยอักษรไทยน้อยแล้วเปรียบเทียบความแตกต่างของการประสมอักษรเป็นคำระหว่างอักษรไทยน้อยและอักษรไทยปัจจุบัน ผลการศึกษาที่ได้จะนำไปใช้ในการออกแบบและพัฒนาตัวแบบสำหรับแปลคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ผลการศึกษารูปได้ ดังนี้

6.1.1.1 นักภาษาโบราณส่วนใหญ่จะมีทักษะในการอ่านอักษรโบราณได้อย่างน้อย 2 อักษร และเข้าใจภาษาที่ใช้บันทึกในเอกสาร โบราณมากกว่า 1 ภาษา ประสบการณ์การปริวรรตเอกสาร โบราณมีมากกว่า 10 ปีขึ้นไป การปริวรรตเอกสารโบราณโดยทั่วไปนักภาษาโบราณจะทำ 2 ขั้นตอน ได้แก่ การทำคำปริวรรต และการทำคำอ่านในแบบภาษาถิ่น หรือแบบแปลเป็นภาษาไทยมาตรฐาน สำหรับการปริวรรตคำราชาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยจะปริวรรตจากอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานในขั้นตอนเดียว โดยคำอ่านบางคำสะกดคำตามแบบพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และบางคำสะกดตามเสียงภาษาไทยถิ่นอีสานโบราณ ซึ่งการสะกดคำกรณีนี้ทำให้พบว่า คำเดียวกันสะกดได้หลายแบบไม่เป็นมาตรฐานเดียวกันในการปริวรรตคำราชาอีสานโบราณ ซึ่งอาจส่งผลให้การตีความไม่ถูกต้องได้

การปริวรรตเอกสารโบราณและคำราชาโบราณอีสานที่เขียนในโบราณ นักภาษาโบราณพบปัญหาคล้ายคลึงกันคือ (1) ปัญหาทางกายภาพของเอกสาร โบราณที่เกิดการเสื่อมสภาพ แตกหัก เปราะเปื้อน ตัวอักษรลบเลือน ทำให้ข้อความไม่สมบูรณ์ และ (2) ปัญหาจากรูปอักษร ภาษา และอักษรวิธีของอักษรไทยโบราณ สิ่งที่เป็นปัญหาในระดับมากคือ อักษรไทยโบราณ ไม่มีวรรณยุกต์ การเขียนคำแบบแยกตัวสะกด หรือการตัดคำในลักษณะที่ทำให้เกิดความไม่แน่ใจในการอ่านและตีความผิดพลาด สำหรับสิ่งที่เป็นปัญหาในระดับน้อย ได้แก่ คำพ้องรูปพ้องเสียง รูปอักษรที่ต่างจากที่เคยเห็น การเขียนผิด ตก หรือ เกิน ทั้งนี้เนื่องจากนักภาษาโบราณส่วนใหญ่มีประสบการณ์สูงในการปริวรรตเอกสาร โบราณ และมีความรู้ในอักษรและภาษาโบราณที่จะปริวรรตอยู่แล้ว สำหรับเนื้อหาหมวดคำราชาที่มีคำศัพท์เฉพาะด้านอยู่เป็นจำนวนมาก จำเป็นต้องอาศัยหมอยาพื้นบ้านอีสานมาช่วยเพื่อให้การปริวรรตมีความถูกต้อง โดยเฉพาะการทำคำอ่าน รูปทักกะที่จำเป็นในการปริวรรตคำราชาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยได้ 5 ด้าน ดังนี้ (1)

รู้จักอักษรไทยน้อยและอักษรธรรมอีสาน (2) รู้และเข้าใจภาษาไทยถิ่นอีสานและบาลี (3) มีความรู้เกี่ยวกับตำรายาโบราณอีสาน (4) มีประสบการณ์ในการปริวรรตอักษรไทยน้อยและอักษรธรรมอีสานมาเป็นเวลานาน และ (5) มีประสบการณ์ในการปริวรรตเอกสารโบราณหลายหมวดหมู่หลายเนื้อหา

6.1.1.2 ตำรายาโบราณอีสานจะบันทึกลงในใบลานขนาดสั้น พื้นที่ด้านซ้ายใช้เขียนเลขหน้า ชื่อตำรับยา หรือชื่อโรค และพื้นที่ด้านขวาซึ่งเป็นพื้นที่ส่วนใหญ่ใช้เขียนตำรับยา การบันทึกจะบันทึกทั้ง 2 หน้า หน้าละ 4 - 6 บรรทัด เขียนต่อเนื่องกันโดยไม่เว้นวรรค และเมื่อเขียนตรงกับรู้อยู่สายสนองจะเว้นข้ามไป ซึ่งบางครั้งทำให้เกิดการแยกพยัญชนะ คำ หรือ พยางค์ ที่ทำให้แบ่งคำอ่านและตีความผิดได้ง่าย รวมถึงการตัดคำขึ้นบรรทัดใหม่และหน้าลานใหม่ด้วย ภาษาที่ใช้บันทึกตำรายาจะบันทึกเป็น 2 ภาษาปะปนกัน ได้แก่ ภาษาไทยถิ่นอีสาน และ ภาษาบาลี เนื่องจากในตำรับยาใช้วิธีรักษาโรคด้วยสมุนไพรและคาถา (บันทึกด้วยภาษาบาลี) หรือใช้ทั้ง 2 วิธีร่วมกัน

6.1.1.3 ลักษณะโครงสร้างการเขียนตำรายาโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อย พบว่า เนื้อหาแบ่งเป็น 2 ประเภท ได้แก่ เนื้อหาประเภทตำรับยา และเนื้อหาอื่นที่ไม่ใช่ตำรับยา ดังนี้ (1) เนื้อหาตำรับยา แบ่งเป็น 3 ประเภท ได้แก่ ตำรับยาที่ใช้เฉพาะสมุนไพรในการรักษา ตำรับยาที่ใช้สมุนไพรร่วมกับพิธีกรรมหรือคาถา และตำรับยาประเภทใช้คาถาในการรักษา ซึ่งตำรับยาแต่ละตำรับมีทั้งชนิดรักษาโรคหรืออาการได้เพียงโรคเดียว อาการเดียว และชนิดที่รักษาได้หลายโรค หลายอาการ บางตำรับมีหลายขนาน แต่ละขนานใช้ตามอาการ ตามเพศ หรือตามอายุของผู้ป่วย คำศัพท์ในกลุ่มนี้ อาทิ ชื่อยา ชื่ออาการป่วย/ชื่อโรค สมุนไพร วิธีเตรียมยา/ปรุงยา วิธีใช้ยา สรรพคุณยา คำรักษา คาถา ผู้ป่วย พิธีกรรมรักษาโรค (2) เนื้อหาอื่นที่ไม่ใช่ตำรับยา เช่น ถกษยามงคล หรือวันเวลาในการเก็บสมุนไพร เป็นต้น คำศัพท์ในกลุ่มนี้ อาทิ ตำรายาโบราณอีสาน วันเวลา วิธีเก็บสมุนไพร ชื่อพิธีกรรม คำกริยาทั่วไป คำนามทั่วไป คำวิเศษณ์ คำคุณศัพท์ คำสันธาน คำบุพบท และคำลงท้ายประโยค สามารถจัดกลุ่มองค์ประกอบเนื้อหาได้ 12 กลุ่มหลัก ได้แก่ (1) ชื่อยา/โรค/อาการ (2) ชื่อสมุนไพร (พืชวัตถุ ธาตุวัตถุ สัตว์วัตถุ กระจายยา) (3) สัดส่วน/ปริมาณที่ใช้ (4) วิธีไปเก็บสมุนไพร (5) วิธีเตรียมยา (6) วิธีใช้ยา (7) ผู้ป่วย (คน สัตว์) (8) วันและเวลา (9) คาถา (10) พิธีกรรม (11) คายหรือคำรักษา และ (12) อื่น ๆ

6.1.1.3 ลักษณะของอักษรไทยน้อย และอักษรวิธีของอักษรไทยน้อย แบ่งประเภทอักษรไทยน้อยเป็น 4 ประเภท ได้แก่ พยัญชนะ สระ ตัวเลข และ เครื่องหมายอื่น ๆ โดยพยัญชนะเดี่ยว มี 26 รูป พยัญชนะประสม (พยัญชนะตัวเต็มและตัวครึ่งเขียนติดกัน) มี 4 รูป พยัญชนะตัวสะกด มี 8 รูป สระ มี 22 รูป ตัวเลข มี 10 รูป การประสมอักษรของอักษรไทยน้อยส่วนมากคล้ายคลึงกับอักษรไทยปัจจุบัน และบางส่วนยืมวิธีประสมอักษรของอักษรธรรมอีสานมาใช้ ไม่นิยมใช้พยัญชนะควบกล้ำ (ร ล ว) และตัวการ์นต์ พยัญชนะควบกล้ำ “ว” ใช้ไม่สม่ำเสมอ

การสะกดคำเดียวกันทำได้หลายแบบ การวางสระบนหรือสระล่างมักเขียนไม่ตรงตำแหน่งปกติ การแก้ไขคำผิดมีรูปแบบไม่ชัดเจน ไม่มีรูปวรรณยุกต์ รูปตัวเลขอักษรไทยน้อยบางรูปมีความคล้ายคลึงกับพยัญชนะ หรือมีรูปเดียวกัน และมีการประสมอักษรด้วยอักษรวิधिพิเศษเฉพาะของแต่ละบุคคล

6.1.1.4 การปริวรรตอักษรไทยน้อยบางอักษรเป็นอักษรไทยปัจจุบันขึ้นอยู่กับบริบทของคำ อีกทั้งการที่อักษรไทยน้อยไม่มีรูปวรรณยุกต์สำหรับกำกับโทนเสียง ทำให้การอ่านอักษรไทยน้อยผู้อ่านต้องอ่านโดยการผันวรรณยุกต์ให้ตรงกับความหมายเอาเอง นอกจากนี้ คำในภาษาไทยถิ่นอีสานยังมีคำที่พ้องเสียงกับภาษาถิ่นอื่นแต่มีความหมายคนละความหมาย อีกทั้งคำหนึ่งคำยังมีความหมายได้มากกว่า 1 ความหมาย ทำให้ผู้ที่ไม่เชี่ยวชาญภาษาไทยถิ่นอีสานแปลความหมายผิดได้ง่าย ซึ่งลักษณะเหล่านี้ส่งผลกระทบต่อการศึกษาปริวรรตคำราชาโบราณอีสานแบบอัตโนมัติ ทั้งการทำคำปริวรรตและคำอ่าน

ผลการศึกษาสรุปได้ว่า ปัญหาการทำคำอ่านภาษาไทยถิ่นอีสานจำแนกได้ 4 ประเด็น คือ (1) อักษรไทยน้อยสะกดคำโดยไม่มีวรรณยุกต์กำกับโทนเสียง (2) ใช้อักษรที่มีเสียงคล้ายคลึงกันสะกดคำเดียวกัน (3) ใช้อักษรธรรมอีสานเขียนปะปนกับอักษรไทยน้อย และ (4) การใช้อักษรวิधिพิเศษเฉพาะบุคคล ส่งผลให้คำหนึ่งคำ สามารถสะกดด้วยอักษรไทยน้อยได้หลายแบบ และทำให้มีคำอ่านได้หลายคำ ดังนั้น การทำคำอ่านที่ถูกต้องจึงต้องอาศัยบริบทที่อยู่รอบ ๆ คำ มาช่วยในการวิเคราะห์

งานวิจัยนี้ สรุปรูปแบบความสัมพันธ์ระหว่างคำปริวรรตของอักษรไทยน้อยและคำอ่านภาษาไทยถิ่นอีสานได้ 4 รูปแบบ ดังนี้ (1) 1:1 หมายถึง คำปริวรรต 1 คำมีคำอ่าน 1 คำอ่าน (2) 1:M หมายถึง คำปริวรรต 1 คำมีคำอ่านได้หลายคำอ่าน (3) M:1 หมายถึง คำปริวรรตหลายคำมีคำอ่านเดียว และ (4) M:N หมายถึง คำปริวรรตแต่ละคำมีคำอ่านได้หลายคำอ่าน และคำอ่านแต่ละคำมีคำปริวรรตได้หลายคำ และได้นำเสนอแนวทางการแก้ปัญหาการแปลงคำปริวรรตเป็นคำอ่านแบบอัตโนมัติสำหรับกรณีความสัมพันธ์แบบ 1:1 และ แบบ M:1 โดยการออกแบบคลังศัพท์หรือพจนานุกรม (Lexicon) ที่เหมาะสม ส่วนความสัมพันธ์แบบ 1:M และ M:N ที่ไม่สามารถแก้ปัญหาการแปลงคำปริวรรตเป็นคำอ่านให้ถูกต้องได้โดยใช้คลังศัพท์เพียงอย่างเดียว แต่ต้องอาศัยบริบทของคำที่อยู่รอบข้างเข้ามาช่วยนั้น ได้ออกแบบออนโทโลยี (Ontology) ซึ่งเป็นเครื่องมือที่สามารถแสดงความสัมพันธ์ที่ซับซ้อนระหว่างคำศัพท์แต่ละกลุ่มมาช่วยในการกำหนดคำอ่านที่เหมาะสม ดังนั้นในการพัฒนาตัวแบบ (Model) เพื่อแปลงคำปริวรรตของอักษรไทยน้อยให้เป็นคำอ่านภาษาไทยถิ่นอีสานและภาษาไทยปัจจุบันครั้งนี้ จึงได้พัฒนาคลังศัพท์และออนโทโลยีหมวดคำราชาโบราณอีสานขึ้นมา เพื่อใช้เป็นเครื่องมือของตัวแบบ

6.1.2 ระยะที่ 2 การออกแบบและพัฒนากรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

กรอบการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ประกอบด้วยองค์ประกอบ 3 องค์ประกอบได้แก่ คลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ออนโทโลยีคำราชาโบราณอีสาน และขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน สรุปผลได้ดังนี้

6.1.2.1 การพัฒนาคังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน มีวัตถุประสงค์เพื่อใช้สำหรับเป็นฐานความรู้สำหรับกระบวนการตัดคำและการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานแบบอัตโนมัติ ประกอบด้วย

1) การศึกษาลักษณะคลังศัพท์ภาษาไทยถิ่นอีสานสำหรับการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานแบบอัตโนมัติ โดยการวิเคราะห์โครงสร้างการอธิบายคำศัพท์ในคลังศัพท์ภาษาไทยถิ่นอีสาน 7 เล่ม และพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 รวมทั้งสิ้น 8 เล่ม กำหนดองค์ประกอบในการวิเคราะห์ได้ 11 องค์ประกอบ ได้แก่ อักษร ภาษามีหมวดคำ มีคำศัพท์ภาษาอีสานโบราณ มีชนิดของคำ มีที่มาของคำหรืออ้างอิง มีตัวอย่างประโยค มีคำตรงข้าม มีคำคล้ายคลึง และมีคำปริวรรต พบว่า โครงสร้างการอธิบายคำศัพท์ของคลังคำภาษาอีสานที่มีในปัจจุบันมีข้อจำกัดในการนำไปใช้แปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานแบบอัตโนมัติ คือ ค้นหาอ่านด้วยคำปริวรรตไม่ได้ คำศัพท์หมวดคำราชาโบราณอีสานมีน้อย และการจัดเก็บคำศัพท์ทั้งหมดอยู่ในรูปแบบหนังสือ

2) การวิเคราะห์และออกแบบโครงสร้างคลังศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ประกอบด้วยเอ็นทิตี 4 เอ็นทิตี ได้แก่ (1) เอ็นทิตีคำศัพท์ (Vocabulary) (2) เอ็นทิตีแหล่งอ้างอิง (Reference) (3) เอ็นทิตีคำปริวรรต (Transliteration) และ (4) เอ็นทิตีคำปริวรรตและคำอ่าน (TransHasVocab) โดยชนิดของคำ (Part-Of-Speech: POS) ที่ใช้ในงานวิจัยนี้ จำแนกเป็น 36 ชนิด ได้แก่ NPRP (วิสามานยนาม) NSMPT (คำนาม หรือ กลุ่มคำ (วลี) ประเภท ชื่อโรค) NP MED (คำนามประเภทชื่อสมุนไพร) NPTBE (คำนามประเภทชนิดของยา) NTME (คำนามประเภทเวลา) NDAY (คำนามประเภทชื่อวัน) NDRT (คำนามประเภทชื่อทิศ) NORG (คำนามประเภทชื่ออวัยวะ) NMTH (คำนามประเภทชื่อเดือนทางจันทรคติ) NCNM (คำนามประเภทจำนวนนับ) NORD (คำนามประเภทคำบอกลำดับ) PDMN (สามานยนาม) PPRS (สรรพนามของคน) PREL (ประพันธสรรพนาม) VERB (คำกริยาทั่วไป ที่ไม่ใช่กริยาในกลุ่มอาการป่วย วิธีเตรียมยา วิธีใช้ยา) VSTMP (คำกริยาอาการป่วย) VMTD (คำกริยาประเภทวิธีเตรียมยา) VDRCT (คำกริยาประเภทวิธีใช้ยา) XVRB (กริยาช่วย) DCNM (คำบอกจำนวน) ADVN (คำกริยาวิเศษณ์รูปปกติ) ADVI (คำกริยาวิเศษณ์รูปคำซ้ำ) ADJN (คำคุณศัพท์รูปปกติ) ADJI (คำคุณศัพท์รูปคำซ้ำ)

CUNT (ชื่อหน่วยนับ) CLTV (ลักษณะนามแบบกลุ่ม) CMTR (ลักษณะนามมาตรวัด) CFQC (คำบอกลักษณะความถี่) CCJT (คำสัญฐาน) PREP (คำบุพบท) EAFF (คำจบประโยค) NEG (คำปฏิเสธ) NEGI (คำปฏิเสธรูปคำซ้ำ) PUNC (เครื่องหมายวรรคตอน) SPELL (คำตา) และ UNK (คำที่ระบุหมวดคำไม่ได้)

3) เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานในคลังศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ผลจากการวิเคราะห์เกณฑ์การสะกดคำศัพท์ภาษาไทยถิ่นอีสานในคลังคำภาษาไทยถิ่นอีสานและตำราภาษาโบราณอีสาน พบว่า มีเกณฑ์การสะกดคำแตกต่างกันขึ้นอยู่กับผู้สร้างคลังศัพท์ ได้แก่ บางคำสะกดคำตามเสียงภาษาไทยถิ่นอีสาน บางคำสะกดคำตรงตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 และบางคำสะกดคำตามคลังคำหรือพจนานุกรมภาษาไทยถิ่นอีสาน ดังนั้น คำภาษาไทยถิ่นอีสานคำเดียวกันจึงอาจมีการสะกดคำต่างกัน และบางคำจะอ่านออกเสียงไม่ตรงกับภาษาไทยถิ่นอีสาน สำหรับงานวิจัยนี้ได้กำหนดเกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานสำหรับคลังศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทย โดยใช้วิธีผสมผสานวิธีการสะกดคำตามแบบภาษาไทยมาตรฐานและการสะกดคำตามเสียงภาษาไทยถิ่นอีสานแบบโบราณและแบบปัจจุบัน ที่เป็นสำเนียงพูดจังหวัดมหาสารคามและกาฬสินธุ์ (ตามสำเนียงของผู้ปริวรรตตำราฉบับที่นำมาศึกษา) สรุปเกณฑ์ที่ใช้สะกดคำอ่านภาษาไทยถิ่นอีสานในงานวิจัยนี้ได้ 2 ประการ คือ

3.1) คำภาษาไทยถิ่นอีสานที่ใช้เกณฑ์สะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ได้แก่ (1) คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้องความหมายกับคำอ่านภาษาไทยมาตรฐาน เช่น กิน เป็น ดอกไม้ เห็ด (2) คำที่มีตัวการันต์ เช่น จันทน์ ประจักษ์ (3) คำที่มีคำปริวรรตถ่ายถอดมาจากพยัญชนะต้น “ย หางยาว” อักษรไทยน้อย เช่น อย่า อยู่ อย่าง อยาก และ (4) คำมีหน่วยเสียงพยัญชนะต้นเป็น ย นาสิก เช่น หมึง หมู้า ไหญ่ ยาก เป็นต้น

3.2) คำภาษาไทยถิ่นอีสานที่ใช้เกณฑ์สะกดคำตามเสียงภาษาไทยถิ่นอีสาน ได้แก่ (1) คำภาษาไทยถิ่นอีสานที่พ้องเสียงพ้องความหมายกับคำภาษาไทยมาตรฐานที่เป็นคำควบกล้ำ ร ล ว เช่น ก้วย เกือ กวง (ไม่มีพยัญชนะควบกล้ำ) (2) คำที่มีหน่วยเสียงพยัญชนะ /ส/ ซึ่งตรงกับเสียงพยัญชนะ /ร/ ในภาษาไทยมาตรฐาน เช่น เสือน ฮ้าง (3) คำที่มีหน่วยเสียงพยัญชนะ /ซ/ ที่ตรงกับหน่วยเสียงพยัญชนะ /ซ/ ในภาษาไทยมาตรฐาน เช่น ซ้าง เซือก (4) คำที่มีคำปริวรรตประสมด้วยพยัญชนะเสียงคล้ายกัน เช่น /ซ/ /ส/ /ท/ หรือประสมด้วยสระที่มีเสียงคล้ายกัน เช่น ชะบับ/ชาบับ/สะบับ สะกดเป็น ชะบับ หรือ เสียบ/เซียบ สะกดเป็น เสียบ หรือ ผักค่านิน/ผักค่านิน สะกดเป็น ผักค่านิน และ (5) คำเฉพาะของภาษาไทยถิ่นอีสาน เช่น กะท้าง (กิ่งก่ายักษ์) ผอง (ดวง) สุก (ถู) เป็นต้น

4) การพัฒนาคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ประกอบด้วย 2 ขั้นตอน ได้แก่ (1) การสร้างคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-ภาษาไทยมาตรฐาน (IsanMedDictionary) สร้างด้วยโปรแกรม SQLiteStudio ตามโครงสร้างของคลังศัพท์ที่ได้ออกแบบไว้ และ (2) การรวบรวมคำศัพท์ จัดเตรียมคำศัพท์ และจัดเก็บคำศัพท์ในฐานข้อมูล ได้แก่ รวบรวมคำศัพท์จากตำรายาโบราณอีสานที่ปริวรรตมาจากตำรายาโบราณอีสานที่ปริวรรตจากอักษรไทยน้อย จำนวน 11 ฉบับ 459 หน้าลาน มาตัดคำด้วยมือ โดยใช้เกณฑ์การตัดคำแบบสั้นและแบบยาวที่สุด เพื่อให้ได้จำนวนคำศัพท์ที่เกี่ยวข้องมากที่สุด เช่น “ฮิมตากะด้าง” (ขอบตากะด้าง) ตัดคำได้คำศัพท์ทั้งหมด 5 แบบ ได้แก่ “ฮิม” “ตา” “กะด้าง” “ฮิมตา” และ “ฮิมตากะด้าง” ผลการดำเนินการขั้นต้นได้คำศัพท์ทั้งสิ้น 25,903 คำ ทำการตรวจแก้การสะกดคำตามเกณฑ์การสะกดคำที่กำหนดเอาไว้ ทำคำปริวรรตโดยการแปลงคำอ่านกลับเป็นคำปริวรรตด้วยเกณฑ์การปริวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบันแบบตรงตัว และตรวจทานคำปริวรรตกับข้อความในเอกสารต้นฉบับ กำกับชนิดคำตามที่ได้ออกแบบไว้ ตัดคำซ้ำ ทำคำอ่านภาษาไทยมาตรฐาน กำหนดคำอธิบาย ได้คำศัพท์ทั้งสิ้น 4,645 คำ คำศัพท์ส่วนใหญ่ร้อยละ 61.53 เป็นคำศัพท์ที่มีความสัมพันธ์แบบ 1:1 อันดับสองร้อยละ 36.10 เป็นแบบ M:1 อันดับสามร้อยละ 1.38 เป็นแบบ 1:M และน้อยที่สุดร้อยละ 0.99 เป็นแบบ M:N นำคำศัพท์ไปตรวจความถูกต้องโดยผู้เชี่ยวชาญที่มีประสบการณ์การปริวรรตตำรายาโบราณอีสานอักษรไทยน้อย แก้ไขตามข้อเสนอแนะ แล้วนำคำศัพท์ไปจัดเก็บในฐานข้อมูล

5) ผลการประเมินประสิทธิภาพของคลังคำโดยนักภาษาโบราณ จำนวน 12 คน ค่าเฉลี่ยทุกรายการอยู่ในระดับดี $\bar{X} = 3.89$ และ $SD = 0.71$ รายการประเมินที่มีค่าเฉลี่ยมากที่สุด อยู่ในระดับดี คือ โครงสร้างคลังคำมีความเหมาะสมสำหรับนำไปใช้แปลงคำปริวรรตเป็นคำอ่านทั้งแบบภาษาไทยถิ่นอีสานและคำอ่านภาษาไทยมาตรฐาน $\bar{X} = 4.08$ และ $SD = 0.67$ และรายการประเมินที่มีค่าเฉลี่ยน้อยที่สุดอยู่ในระดับดี คือ จำนวนคำศัพท์มีเพียงพอสำหรับใช้ปริวรรตตำรายาโบราณอีสาน $\bar{X} = 3.55$ และ $SD = 0.82$ และผลการประเมินความเหมาะสมของเกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสานค่าเฉลี่ยทุกด้านอยู่ในระดับดี $\bar{X} = 3.91$ และ $SD = 0.72$ โดยเกณฑ์สะกดคำอ่านภาษาไทยถิ่นอีสานที่มีค่าเฉลี่ยสูงสุด อยู่ในระดับดี คือ คำปริวรรตที่มีคำอ่านเป็นคำภาษาไทยถิ่นอีสานที่สะกดคำตามคลังคำภาษาไทยถิ่นอีสาน เช่น “คันคาก” (คางคก) “กะท้าง” (กึ่งก่าขัณฑ์) เป็นต้น $\bar{X} = 4.08$ และ $SD = 0.67$ และรายการประเมินที่มีค่าเฉลี่ยน้อยที่สุดมี 2 รายการเท่ากัน อยู่ในระดับดี ได้แก่ คำอ่านภาษาไทยถิ่นอีสานที่คำปริวรรตประสมคำด้วยพยัญชนะต้น /ช//ส//ท/ และสระที่มีเสียงคล้ายกัน (สระเสียงสั้น/ยาว) การสะกดคำอ่านภาษาไทยถิ่นอีสานใช้พยัญชนะและสระตรงกับคำภาษาไทยถิ่นอีสานสมัยปัจจุบัน ยกตัวอย่าง เช่น “ซาบับ/ชะบับ” (ฉบับ) เขียนสะกดเป็น “ฉบับ” “ชะหน้อย/สน้อย” เขียนสะกดเป็น “สน้อย” และคำภาษาไทยถิ่น

อีสานที่ คำปริวรรตประสมคำด้วยพยัญชนะและสระเสียงคล้ายกัน และเป็นคำที่มีความหมายเดียวกัน คำอ่านสะกดคำตามเสียงภาษาไทยถิ่นอีสานปัจจุบัน “กิน/กิน” เขียนสะกดเป็น “กิน” “จิง/จิง” เขียนสะกดเป็น “จิง/จิง” ค่าเฉลี่ย $\bar{X} = 3.75$ SD = 0.62 และ 0.75 ตามลำดับ

6.1.2.2 การพัฒนาออนโทโลยีตำรายาโบราณอีสาน

งานวิจัยนี้ได้ทำการพัฒนาออนโทโลยีตำรายาโบราณอีสานเพื่อแก้ปัญหากรณีคำปริวรรตหนึ่งคำมีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำ (1:M) หรือ คำปริวรรตแต่ละคำมีคำอ่านภาษาไทยถิ่นอีสานได้หลายคำอ่าน และคำอ่านภาษาไทยถิ่นอีสานแต่ละคำอ่านนั้นมีคำปริวรรตได้หลายคำ (M:N) เนื่องจากเมื่อนำคำปริวรรตไปค้นหาคำอ่านจากคลังศัพท์จะได้คำอ่านหลายคำอ่าน และการเลือกคำอ่านที่ถูกต้องจะต้องอาศัยคำบริบทที่อยู่รอบข้างคำนั้นมาช่วยออนโทโลยีตำรายาโบราณอีสานที่พัฒนาขึ้นจะจัดเก็บคำศัพท์ภาษาไทยถิ่นอีสานและคำบริบทที่อยู่ติดกันด้านซ้ายและขวาที่มีความสัมพันธ์สอดคล้องกัน โดยขอบเขตความรู้ในออนโทโลยี แบ่งเป็น 2 กลุ่ม ได้แก่ เนื้อหาด้านตำรับยา และ เนื้อหาอื่น ๆ ที่ไม่ใช่ตำรับยา โครงสร้างของออนโทโลยีประกอบด้วย 20 คลาส ตามอนุกรมวิธานคำศัพท์ในตำรายาโบราณอีสาน ได้แก่

- 1) คลาสชื่อยา (Medicine) ประกอบด้วย 2 คลาสย่อย คือ ยารักษามนุษย์ และยารักษาสัตว์
- 2) คลาสรสยา (Taste)
- 3) คลาสอาการป่วย/ชื่อโรค (Symptoms)
- 4) คลาสเภสัชวัตถุ (FormulationMaterial) ประกอบด้วย 4 คลาสย่อย คือ พืชวัตถุ สัตว์วัตถุ ธาตุวัตถุ และกระสายยา
- 5) คลาสวิธีเตรียมยา/ปรุงยา (Pharmacy) ประกอบด้วย 3 คลาสย่อย คือ วิธีเตรียมยา/ปรุงยา อุปกรณ์ปรุงยา และรูปแบบผลิตภัณฑ์
- 6) คลาสปริมาณ (Quantity) ประกอบด้วย 2 คลาสย่อย คือ จำนวนตัวเลข และคำบอกปริมาณ
- 7) คลาสหน่วยวัดและชั่งตวง (Measurement)
- 8) คลาสวิธีใช้ยา (Treatment)
- 9) คลาสค่ารักษา (MedicineFees)
- 10) คลาสผู้ป่วย (Patient) ประกอบด้วย 2 คลาสย่อย คือ มนุษย์ และสัตว์
- 11) คลาสพิธีกรรม (Rite) ประกอบด้วย 3 คลาสย่อย คือ กิจกรรมอุปกรณ์/เครื่องประกอบพิธีกรรม และคาถา
- 12) คลาสวันเวลา (DateTime) ประกอบด้วย 3 คลาสย่อย คือ คำเรียกเวลา คำเรียกวัน และคำเรียกเดือน

- 13) คลาสคำกริยาอื่น ๆ (Verb) ที่ไม่อยู่ในหมวดคำรับยา
- 14) คลาสชื่อสัตว์ (Animal) ประกอบด้วย 3 คลาสย่อย คือ ชื่อสัตว์ ส่วนประกอบ/อวัยวะสัตว์ และคำเรียกเพศ
- 15) คำนามอื่น ๆ (Noun) ประกอบด้วย 4 คลาสย่อย คือ นามทั่วไป สรรพนาม ชื่อทิส และหน่วยนับ
- 16) คำวิเศษณ์ (Adverb) ประกอบด้วย 2 คลาสย่อย คือ คำวิเศษณ์หน้ากริยา และคำวิเศษณ์หลังกริยา
- 17) คำคุณศัพท์ (Adjective) ประกอบด้วย 2 คลาสย่อย คือ คำบ่งชี้ และคำบอกคุณลักษณะ
- 18) คำสันธาน (Conjunction)
- 19) คำบุพบท (Preposition:)
- 20) คำลงท้ายประโยค (End)

Data Properties ของแต่ละคลาส ประกอบด้วย คำอ่านภาษาไทยถิ่นอีสาน คำอ่านภาษาไทยมาตรฐาน และรหัสคำศัพท์ และ Object Properties ประกอบด้วย คำศัพท์ที่อยู่ในตำแหน่งติดกันข้างขวาและข้างซ้าย โดยหลักการสอบถามคำอ่านภาษาไทยถิ่นอีสานที่ถูกต้องด้วยออนโทโลยี ใช้วิธีนำเอาคำอ่านภาษาไทยถิ่นอีสานและคำบริบทที่อยู่ติดกับคำอ่านทั้งด้านซ้ายและด้านขวา ไปค้นหาว่ามีคำศัพท์ใดในออนโทโลยีที่มีความสัมพันธ์ตรงกับคำอ่านที่ค้นหาหรือไม่ ถ้าพบแสดงว่าคำอ่านดังกล่าวมีความเหมาะสมที่จะใช้เป็นคำอ่านภาษาไทยถิ่นอีสานได้ ซึ่งผลการประเมินประสิทธิภาพการใช้ออนโทโลยีคำราชาโบราณอีสานสำหรับทำคำอ่านภาษาไทยถิ่นอีสานที่ความสัมพันธ์ระหว่างคำปริวรรตและคำอ่านเป็นแบบ 1:M และ M:N จำนวน 1,758 คำ พบว่าประสิทธิภาพความถูกต้องตรงกับคำเฉลยอยู่ในระดับร้อยละ 75.88 โดยความผิดพลาดเกิดจาก 2 กรณี คือ แปลงคำอ่านผิดเป็นคำอื่น และเป็นคำที่ไม่รู้จัก หรือไม่มีคำอ่านและคำบริบทที่สอบถามในออนโทโลยี อยู่ในระดับร้อยละ 22.53 และร้อยละ 1.59 ตามลำดับ

6.1.2.3 ขั้นตอนวิธีแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน แบ่งเป็น 3 ขั้นตอน ได้แก่

- 1) การตัดคำ คือการนำคำรับยาแบบคำปริวรรตมาตัดเป็นคำด้วยโปรแกรมตัดคำ (Word Tokenizer) ของ PythaiNLP ที่ใช้คลังศัพท์คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-ภาษาไทยมาตรฐาน เป็นฐานความรู้สำหรับตัดคำ พบปัญหา 2 ประการ ได้แก่ (1) การตัดคำผิดพลาด เนื่องจากคำศัพท์ในคลังศัพท์ไม่ครอบคลุมคำศัพท์ในข้อมูลชุดทดสอบ ซึ่งกรณีนี้จะส่งผลกระทบต่อเนื่องไปยังขั้นตอนการทำคำอ่านภาษาไทยถิ่นอีสานเกิดความผิดพลาดตามไปด้วย และ (2)

คำปวิวรรต 1 คำ ตัดคำได้หลายแบบขึ้นอยู่กับบริบท เช่น “นำทา” ตัดได้ 2 แบบคือ แบบที่ 1 “นำ” (นำ/น้ำ) และ “ทา” แบบที่ 2 “นำทา” (น้ำท่า) เป็นต้น เพื่อไม่ให้ผลการตัดคำผิดพลาดส่งผลกระทบต่อการทำคำอ่าน ผู้วิจัยจึงดำเนินการแก้ไขการตัดคำผิดพลาดโดยการแก้ไขให้ถูกต้องด้วยมือ

2) การค้นหาคำอ่านในคลังศัพท์ คำปวิวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ได้แก่ นำคำปวิวรรตที่ตัดคำเรียบร้อยแล้วแต่ละคำไปค้นหาคำอ่านภาษาไทยถิ่นอีสานที่เป็นไปได้ทั้งหมดในฐานข้อมูลคลังศัพท์ คำปวิวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ผลลัพธ์ที่ได้มี 3 รูปแบบ ได้แก่ (1) ไม่พบคำอ่านภาษาไทยถิ่นอีสาน (จำนวนคำอ่านภาษาไทยถิ่นอีสาน = 0) กรณีนี้จะใช้คำปวิวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน (คำอ่าน = คำปวิวรรต) (2) พบคำอ่านภาษาไทยถิ่นอีสาน 1 คำอ่าน คำอ่านนั้นจะถูกเลือกไปใช้เป็นคำอ่าน และ (3) พบคำอ่านภาษาไทยถิ่นอีสานมากกว่า 1 คำอ่าน คำอ่านทั้งหมดจะถูกนำไปหาคำอ่านที่เหมาะสมด้วยออนโทโลยีในลำดับต่อไป

3) การสอบถามคำอ่านที่เหมาะสมในออนโทโลยีคำราชาโบราณอีสาน ได้แก่ การนำคำอ่านทั้งหมดพร้อมคำบริบทที่อยู่ติดกับคำอ่านด้านซ้ายและขวา ข้างละ 1 คำ ไปสอบถามในออนโทโลยีคำราชาโบราณอีสาน ผลลัพธ์ที่ได้มี 2 รูปแบบ ได้แก่ ไม่พบคำอ่าน (จำนวนผลลัพธ์ = 0) ในกรณีนี้จะใช้คำปวิวรรตเป็นคำอ่าน และ พบคำอ่าน (จำนวนผลลัพธ์มากกว่า 1 คำ) ในกรณีนี้คำอ่านนี้จะถูกเลือกไปเป็นคำอ่าน

6.1.3 ระยะที่ 3 การประเมินประสิทธิผลของกรอบแนวคิดในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน สรุปได้ดังนี้

6.1.3.1 ความถูกต้องแม่นยำของกรอบแนวคิด (Accuracy) วิธีประเมิน ได้แก่ นำเอาข้อมูลคำปวิวรรตคำราชาโบราณอีสานโบราณจำนวน 807 คำรับ (7,929 คำ) ไปประมวลผลด้วยโปรแกรมแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน พบว่ามีระดับความถูกต้อง 6,965 คำ หรือร้อยละ 87.84 โดยความสัมพันธ์แบบ 1:1 มีความถูกต้องแม่นยำสูงสุด อยู่ในระดับ ร้อยละ 93.29 อันดับสอง แบบ M:1 ร้อยละ 86.20 อันดับสาม แบบ 1:M ร้อยละ 81.33 และอันดับสุดท้าย คือ แบบ M:N ร้อยละ 74.01 ความผิดพลาดของตัวแบบเกิดใน 2 กรณี ได้แก่ (1) คำศัพท์ที่ค้นหาเป็น “คำที่ไม่รู้จัก” เกิดจากคำปวิวรรตนั้นไม่มีคำอ่านในคลังศัพท์ หรือมีคำอ่านในคลังศัพท์แต่ไม่มีในออนโทโลยีคำราชาโบราณอีสาน และ (2) แปลงคำผิด หรือ คำอ่านภาษาไทยถิ่นอีสาน ไม่ถูกต้อง ซึ่งมีสาเหตุจาก คำปวิวรรตเป็นคำประเภท 1:M แต่มีคำอ่านในคลังศัพท์เพียง 1 คำ ทำให้ตัวเลือกในการนำไปใช้เป็นคำอ่านมีเพียงตัวเลือกเดียว ที่ทำให้ในบางบริบทคำอ่านดังกล่าวเป็นคำอ่านที่ไม่สอดคล้องกับบริบท และคำบางคำสามารถตัดคำได้หลายแบบขึ้นอยู่กับบริบท ซึ่งการตัดคำผิดจะส่งผลกระทบต่อแปลงคำปวิวรรตเป็นคำอ่านผิดพลาดตามไปด้วย

6.1.3.2 การประเมินประสิทธิภาพการทำงานในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ใช้วิธีประเมินประสิทธิภาพโดยการสร้างระบบต้นแบบ ซึ่งวิเคราะห์ ออกแบบ และพัฒนาระบบตามกระบวนการปริวรรตเอกสารโบราณของนักภาษาโบราณ และขั้นตอนวิธีที่ผู้วิจัยพัฒนาขึ้นมา ข้อมูลที่นำเข้าสู่ระบบเพื่อทำการทดสอบ ได้แก่ เก็บรวบรวมข้อมูลด้วยแบบประเมินประสิทธิภาพตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ผู้ประเมินระบบแบ่งเป็น 2 กลุ่ม ได้แก่ นักภาษาโบราณจำนวน 12 คน และผู้ใช้ประโยชน์จากตำรายาโบราณอีสานจำนวน 27 คน สรุปผลได้ดังนี้

1) ผลการพัฒนาระบบต้นแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน ประกอบด้วยเมนูหลัก ได้แก่ เมนูปริวรรตตำรายาฉบับใหม่ ใช้สำหรับแปลงคำปริวรรตเป็นคำอ่าน เมนูจัดการข้อมูลตำรายา ใช้สำหรับแก้ไขปรับปรุงตำรายาที่ปริวรรตเป็นคำอ่านแล้ว เมนูอักษรไทยน้อย ใช้สำหรับอธิบายลักษณะอักษรไทยน้อยที่ใช้เขียนตำรายาโบราณอีสาน เมนูอักษรวิธีอักษรไทยน้อย ใช้สำหรับอธิบายอักษรวิธีการประสมอักษรไทยน้อยเป็นคำอ่านในตำรายาโบราณอีสาน และเมนูคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน ใช้สำหรับค้นหาคำอ่านภาษาไทยถิ่นอีสานจากคำปริวรรต

2) ผลการประเมินประสิทธิภาพตัวแบบ สรุปได้ว่า ผู้ประเมินทั้งสองกลุ่มมีค่าเฉลี่ยทุกรายการอยู่ในระดับดี $\bar{X} = 4.08$ และ $SD = 0.73$ และรายการที่มีค่าเฉลี่ยสูงสุดเท่ากันอยู่ในระดับดี มี 2 รายการ ได้แก่ ระบบต้นแบบฯ ช่วยให้การสะกดคำอ่านภาษาไทยถิ่นอีสานในตำรายาอีสานเป็นมาตรฐานเดียวกันมากขึ้น และระบบต้นแบบฯ ช่วยให้ผู้ที่รู้จักอักษรไทยน้อยแต่ไม่รู้หรือไม่เชี่ยวชาญภาษาไทยถิ่นอีสาน และไม่เชี่ยวชาญเนื้อหาตำรายาอีสาน สามารถอ่านตำรายาโบราณอีสานอักษรไทยน้อยได้ถูกต้องมากขึ้น $\bar{X} = 4.21$ และ $SD = 0.66, 0.73$ ตามลำดับ รายการที่มีค่าเฉลี่ยต่ำสุด คือ ความเร็วในการประมวลผลอยู่ในระดับที่ยอมรับได้ อยู่ในระดับดี $\bar{X} = 3.64$ และ $SD = 0.84$ เมื่อเปรียบเทียบค่าเฉลี่ยทุกรายการระหว่างกลุ่มตัวอย่าง 2 กลุ่ม พบว่า กลุ่มผู้ใช้ประโยชน์จากตำรายาโบราณอีสาน ($\bar{X} = 4.18$ และ $SD = 0.73$) ให้ค่าเฉลี่ยทุกรายสูงกว่านักภาษาโบราณ ($\bar{X} = 3.85$ และ $SD = 0.67$)

6.2 อภิปรายผลการวิจัย

6.2.1 กรอบแนวคิดในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

กรอบแนวคิดในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานให้มีความหมายถูกต้องตามบริบทของตำรายาอีสานโบราณ โดยใช้คลังศัพท์และออนโทโลยีเป็นฐานความรู้ที่แน่น ควรมีองค์ประกอบ และกระบวนการ ดังต่อไปนี้

6.2.1.1 องค์ประกอบของกรอบการทำงาน ประกอบด้วย 3 องค์ประกอบหลัก คือ คลังศัพท์คำปวิวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน หมวดยาโบราณอีสาน ที่มีคำหลักเป็นคำปวิวรรตของอักษรไทยน้อย และออนโทโลยีตำรายาโบราณอีสาน โดยใช้เป็นฐานความรู้สำหรับการตัดคำและการทำคำอ่าน พบว่า ระดับความถูกต้องของกระบวนการทั้งสองจะขึ้นอยู่กับจำนวนคำศัพท์ในคลังศัพท์ กล่าวคือ ยังมีคำศัพท์ในคลังศัพท์มากยิ่งขึ้นเพิ่มความถูกต้องในการแปลงคำปวิวรรตเป็นคำอ่านให้เพิ่มสูงขึ้น อย่างไรก็ตาม สำหรับกรณีคำปวิวรรตประเภท 1:M และ M:N นั้น จะไม่สามารถใช้คลังศัพท์ที่มีขนาดใหญ่เพียงอย่างเดียวได้ เนื่องจากคำปวิวรรตประเภทนี้ต้องอาศัยบริบทที่อยู่รอบ ๆ มาช่วยวิเคราะห์หรือเลือกคำอ่านที่ถูกต้อง การแก้ปัญหาดังกล่าวด้วยออนโทโลยีช่วยให้ประสิทธิภาพของตัวแบบโดยรวมเพิ่มสูงขึ้น เมื่อเปรียบเทียบกับการใช้คลังศัพท์ในการแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านเพียงอย่างเดียว โดยเพิ่มจากเดิม ร้อยละ 70.02 เป็นร้อยละ 87.84

6.2.1.2 กระบวนการของตัวแบบในงานวิจัยนี้ได้ปรับใช้ขั้นตอนการประมวลผลภาษาธรรมชาติ (Ghosh & Gunning, 2019) กระบวนการแปลภาษาด้วยเครื่อง แบบแปลทางตรง (Direct Transition) (Abdelaal, & Alazzawie, 2020) และหลักการปวิวรรต (Transliteration) เอกสารโบราณ (พิมพ์พรรณ ไพบูลหวังเจริญ, 2542) ประกอบด้วย 2 ขั้นตอนหลัก คือ (1) การตัดคำ และ (2) การทำคำอ่าน ซึ่งสอดคล้องกับลักษณะการเขียนตำรายาโบราณอีสานที่เขียนต่อเนื่องกันไปเรื่อย ๆ โดยไม่มีการเว้นวรรค และการทำคำอ่านจำเป็นต้องนำคำปวิวรรตมาตัดคำก่อนจึงจะทำคำอ่านได้ จากนั้น คำอ่านที่ได้จะถูกนำมาจัดเรียงเป็นประโยคใหม่ในตำแหน่งเดิมโดยไม่ต้องใช้กฎไวยากรณ์ของภาษาเป้าหมาย เนื่องจากภาษาต้นฉบับ (ภาษาที่ใช้บันทึกตำรายาโบราณอีสานด้วยอักษรไทยน้อย) และภาษาเป้าหมาย (คำอ่านตำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยปัจจุบัน) เป็นภาษาเดียวกัน คือ ภาษาอีสาน กระบวนการที่นำเสนอจึงมีความเหมาะสมกับการแปลงคำปวิวรรตเป็นคำอ่าน ดังจะเห็นได้จากผลการประเมินระบบต้นแบบแปลงคำปวิวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานที่พัฒนาจากตัวแบบที่ได้จากการวิจัยนี้ ซึ่งพบว่า นักภาษาโบราณและผู้ใช้ประโยชน์จากตำรายาโบราณอีสานให้คะแนนความเหมาะสมทุกด้านอยู่ในระดับดี และระยะเวลาการประมวลผลโดยรวมทุกขั้นตอนยังอยู่ในระยะเวลาที่ผู้ใช้รับได้ เมื่อเทียบกับการปวิวรรตเอกสาร

โบราณด้วยมือทุกขั้นตอนโดยมนุษย์ ที่ใช้เวลามากกว่าตั้งแต่เป็นชั่วโมงจนถึงเป็นวันขึ้นอยู่กับความเชี่ยวชาญ รวมถึงระบบต้นแบบ ๆ มีการจัดเก็บข้อมูลคำปริวรรตและคำอ่านไว้ในฐานข้อมูล ทำให้การจัดเก็บข้อมูลเป็นระบบระเบียบมากขึ้น

กระบวนการตัดคำของกรอบการทำงาน ๆ นี้ ใช้โปรแกรมตัดคำ (Word Tokenizer) ของ PythaiNLP และใช้คลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-ภาษาไทยมาตรฐาน แทนคลังศัพท์เดิมของ PythaiNLP ซึ่งไม่สามารถตัดคำปริวรรตได้อย่างมีประสิทธิภาพ พบว่า ขั้นตอนการตัดคำปริวรรตพบปัญหาคล้ายคลึงกับการตัดคำในภาษาไทยมาตรฐาน และภาษาอื่น ๆ ที่เขียนคำติดกันโดยไม่เว้นวรรค คือ “การตัดคำผิด” (เขียนสะกดผิด) ทำให้หาความหมายของคำไม่ได้ และ “การตัดคำถูก” แต่มีความหมายไม่ตรงบริบท

การตัดคำผิดโดยมากมักเกิดกับคำประสมที่เกิดจากคำโดดหลายคำประสมกัน ซึ่งคำศัพท์ในตำรับยาโบราณอีสานจะมีคำประสมหลายพยางค์ (4 - 7 พยางค์) อยู่ค่อนข้างมาก และคำเดียวกันยังเขียนได้หลายแบบ ทั้งแบบใช้พยัญชนะ สระ ต่างกัน หรือตำแหน่งคำสลับกัน โดยเฉพาะชื่อโรคหรือชื่อเรียกอาการ เช่น “ปวดแข็งปวดขา” หรือ “ปวดขาปวดแข็ง” “เมื่อยดินเมื่อยมือ” หรือ “เมื่อยมือเมื่อยดิน” หรือ แบบคำเดียวกันแต่จำนวนพยางค์สั้นยาวไม่เท่ากัน พบในการเขียนชื่อสมุนไพร เช่น “ใบต้นเค็ดคดหมา” “ใบคดหมา” “ใบเค็ดคดหมา” หรือ “เป็กหมากล่า ตาควย” “เป็กหมากล่าตาควย” เป็นต้น นอกจากนี้ ยังพบกรณีที่กำลังคำหนึ่งคำสามารถตัดคำได้หลายแบบขึ้นอยู่กับบริบท เช่น “นำทา” หรือ “นำฝน” ที่ตัดคำได้ 2 แบบ ได้แก่ แบบคำโดด และแบบคำเดียว 2 พยางค์ ดังนี้ นำ, ทา, นำทา (นำ, น้ำ, น้ำทา) หรือ นำ, ฝน, นำฝน (นำ, น้ำ, น้ำฝน) เป็นต้น ซึ่ง การตัดคำแบบยาวที่สุดจะได้คำประสมที่คำอ่านมีความหมายไม่ตรงบริบท หรือ หากเป็นคำที่ไม่มีในคลังศัพท์ก็จะกลายเป็นคำผิด ซึ่งปัญหา “การตัดคำปริวรรตผิด” จะส่งผลกระทบต่อความถูกต้องในการทำคำอ่านโดยตรง อย่างไรก็ตาม ในกระบวนการสร้างคลังศัพท์จะใช้ทั้งการตัดคำแบบยาวที่สุดและการตัดคำแบบสั้นที่สุดเพื่อให้ได้คำศัพท์ที่รองรับความหลากหลายของรูปแบบการสะกดคำ แต่ยังคงไม่สามารถแก้ปัญหาความผิดพลาดในกระบวนการตัดคำและการทำคำอ่านในกรณีนี้ได้ทั้งหมด

6.2.2 คลังศัพท์

การออกแบบและพัฒนาคลังศัพท์ในงานวิจัยนี้ ปรับใช้ขั้นตอนการพัฒนาฐานข้อมูลผสานวิธีกับขั้นตอนการพัฒนาคลังศัพท์แบบใช้คลังคำเป็นฐาน (วิโรจน์ อรุณมานกุล, 2553) ผลการวิจัยได้คลังศัพท์ที่จัดเก็บในระบบฐานข้อมูล ชื่อว่า “คลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน” คำศัพท์ในคลังศัพท์เป็นคำศัพท์หมวดตำรายาโบราณอีสาน มีคำศัพท์หลักเป็นคำปริวรรตของอักษรไทยน้อย พบว่า การออกแบบฐานข้อมูล

คลังศัพท์ ซึ่งประกอบด้วย 4 เอ็นทิตี ได้แก่ คำศัพท์ (Vocabulary) เอกสารอ้างอิง (References) คำปริวรรต (Transliteration) และคำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน (TransHasVocab) ช่วยให้คำศัพท์ในฐานข้อมูลไม่เกิดความซ้ำซ้อนและขัดแย้งกัน เนื่องจากแต่ละเอ็นทิตีถูกปรับให้อยู่ในรูปแบบปกติ (Normalization Form) แล้ว สำหรับผลการประเมินประสิทธิภาพของคลังศัพท์โดยการนำไปทดสอบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานได้ถูกต้องอยู่ในระดับค่อนข้างสูง (ร้อยละ 71.02) เนื่องจากคำศัพท์ที่ใช้เขียนตำราภาษาโบราณอีสานส่วนใหญ่มีการใช้คำซ้ำกันอยู่มาก ดังจะเห็นได้จากขั้นตอนการรวบรวมคำศัพท์จากตำราภาษาโบราณอีสานที่ปริวรรตมาจากอักษรไทยน้อย ขนาด 25,903 คำ หลังตัดคำซ้ำแล้วได้คลังศัพท์ขนาด 4,645 คำ ลดลงร้อยละ 82.07 และความสัมพันธ์ของคำปริวรรตและคำอ่านส่วนใหญ่ร้อยละ 97.63 เป็นความสัมพันธ์แบบ 1:1 และ M:1

ผลการประเมินประสิทธิภาพความถูกต้องในการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานด้วยคลังศัพท์เพียงองค์ประกอบเดียว เมื่อพิจารณาด้านความผิดพลาด (ร้อยละ 28.98) พบว่า ส่วนใหญ่เป็นความผิดพลาดประเภท “คำไม่รู้จัก” เนื่องจาก ไม่พบคำอ่าน และ “แปลงคำอ่านผิด” เนื่องจากคำปริวรรตมีคำอ่านหลายคำอ่าน ทำให้ไม่สามารถเจาะจงคำอ่านที่ถูกต้องได้ หรือมีคำอ่านในคลังศัพท์แต่อ่านแปลไม่ได้ความหมาย (คำอ่านไม่สอดคล้องกับบริบท) ทั้งนี้ สาเหตุเกิดจากจำนวนคำศัพท์ในคลังศัพท์มีค่อนข้างน้อย (4,645 คำ) ทำให้ไม่ครอบคลุมคำศัพท์ที่ใช้เขียนตำรา เพื่อลดความผิดพลาดประเภทนี้ จึงควรให้ความสำคัญกับขนาดของคลังข้อมูลที่จะมาใช้สร้างคลังศัพท์ ซึ่งขนาดคลังข้อความที่แนะนำคือ ตั้งแต่ 1 แสนคำขึ้นไป (วิโรจน์ อรุณมานะกุล, 2533) เพื่อให้ได้คำศัพท์ที่ครอบคลุมคำศัพท์มากที่สุด จะช่วยเพิ่มความถูกต้องให้สูงเพิ่มมากขึ้น อย่างไรก็ดี สำหรับกรณีที่คำปริวรรตประเภท 1:M และ M:N นั้น ยังไม่สามารถใช้คลังศัพท์ที่มีขนาดใหญ่เพียงอย่างเดียวได้ แต่ต้องอาศัยคำบริบทที่อยู่รอบ ๆ มาช่วยวิเคราะห์คำอ่านที่ถูกต้องร่วมด้วย

การพัฒนาคลังศัพท์โดยมีคำศัพท์หลักเป็นคำปริวรรต พบว่า มีข้อดีสำคัญคือ ช่วยลดข้อจำกัดด้านแบบอักษร (Font) ที่ใช้พิมพ์คำศัพท์หลัก เนื่องจากการสร้างคลังศัพท์ อักษรไทยน้อย-คำอ่านภาษาไทยถิ่นอีสาน (อักษรไทยมาตรฐาน) นั้น คำศัพท์หลักจะต้องพิมพ์ด้วยแบบอักษรไทยน้อย ซึ่งปัจจุบันยังไม่มีรหัสยูนิโคด (Unicode) และการใช้คลังศัพท์อักษรโบราณลักษณะนี้จะใช้ทำคำอ่านได้เฉพาะระบบปริวรรตเอกสาร โบราณด้วยเครื่องที่เป็นอักษรไทยน้อยเท่านั้น ในขณะที่คลังศัพท์ที่มีคำศัพท์หลักเป็นคำปริวรรตอักษรไทยปัจจุบัน สามารถนำมาใช้แปลงคำปริวรรตของอักษรธรรมอีสานซึ่งเป็นอักษรโบราณที่คนไทยถิ่นอีสานนิยมนำมาบันทึกตำราได้อีกอักษรหนึ่ง โดยไม่จำเป็นต้องปรับเปลี่ยนโครงสร้างฐานข้อมูลและชุดคำศัพท์ในคลังศัพท์แต่อย่างใด

การพัฒนาคลังศัพท์นี้ ใช้วิธีรวบรวมคำศัพท์จากตำราयीसानที่ปริวรรตมาจากตำราयीसानโบราณที่บันทึกด้วยอักษรไทยน้อย พบว่ามีข้อดี คือ ทำให้ได้คำศัพท์ตรงกับวัตถุประสงค์ในการนำคลังศัพท์ไปใช้งาน และใช้เวลาสร้างน้อยกว่าวิธีรวบรวมคำศัพท์จากคลังศัพท์หรือพจนานุกรมที่มีผู้รวบรวมเอาไว้แล้ว เนื่องจากไม่ต้องเสียเวลาในการค้นหาและเลือกคำศัพท์จากคลังศัพท์หลายแหล่ง ซึ่งส่วนใหญ่เป็นคำศัพท์หมวดทั่วไป ขณะที่คำศัพท์หมวดตำราयीसानโบราณมีน้อย และการตัดคำด้วยมือมีความผิดพลาดต่ำกว่าการตัดคำด้วยโปรแกรมตัดคำภาษาไทย (Thai Word Segmentation) เช่น PythaiNIP TLexPlus หรือ Lexto เนื่องจากโปรแกรมเหล่านี้ไม่ได้สร้างจากฐานความรู้ทางภาษาที่เป็นภาษาไทยถิ่นอีสานโบราณ เนื้อหาหมวดตำราयीसानโดยเฉพาะ เมื่อนำมาใช้ตัดคำปริวรรตหมวดตำราयीसानโบราณทำให้เกิดความผิดพลาดสูง อย่างไรก็ตาม การตัดคำด้วยมือจะมีความรวดเร็วและถูกต้องจะขึ้นอยู่กับความเชี่ยวชาญด้านภาษาศาสตร์ ภาษาไทยถิ่นอีสาน และความรู้ด้านตำราयीसानโบราณของผู้ดำเนินการเป็นหลัก

การกำกับชนิดของคำในคลังศัพท์ การวิจัยนี้ แบ่งชนิดของคำ (POS) เป็น 36 ชนิด โดยมีวัตถุประสงค์เพื่อใช้อธิบายทั้งหน้าที่ของคำตามวากยสัมพันธ์ของประโยคและหมวดคำศัพท์ตามโครงสร้างการเขียนตำราयीसानโบราณอีสานไปพร้อมกัน ซึ่งจะช่วยลดความกำกวมในคำพ้องรูปพ้องเสียงได้ เช่น “เขา” ชนิดของคำคือ NP MED หมายถึง คำนามประเภทชื่อสมุนไพรรหรือชนิดของคำ PPRS หมายถึง คำสรรพนามหมวดคำนามของมนุษย์ เป็นต้น

การทำคำปริวรรตของอักษรไทยน้อยโดยใช้วิธีแปลงคำอ่านกลับเป็นคำปริวรรตพบว่า ช่วยให้เห็นตอนการเตรียมคำศัพท์ทำได้ง่ายมากกว่าการทำคำปริวรรตด้วยวิธีถอดรูปอักษรไทยน้อยจากเอกสารต้นฉบับโดยตรง เนื่องจากการใช้เกณฑ์การปริวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบันมาใช้แปลงคำอ่านกลับเป็นคำปริวรรต ทำให้ผู้ที่มีความรู้ด้านการอ่านอักษรไทยน้อยในระดับพื้นฐานสามารถทำคำปริวรรตได้

เกณฑ์การสะกดคำอ่านภาษาไทยถิ่นอีสาน ใช้เกณฑ์การสะกดคำตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 สำหรับคำพ้องเสียงพ้องความหมายกับคำภาษาไทยมาตรฐาน ยกเว้นคำควบกล้ำและคำภาษาไทยถิ่นอีสานใช้เกณฑ์การสะกดคำตามเสียงภาษาไทยถิ่นอีสานพบว่า ผลการประเมินความเหมาะสมของเกณฑ์การสะกดคำศัพท์ในคลังศัพท์อยู่ในระดับดี เนื่องจาก ช่วยให้อ่านง่าย และตรงกับความต้องการของผู้ใช้ ที่ต้องการให้การปริวรรตเอกสารโบราณอนุรักษ์รูปแบบการใช้คำตามแบบโบราณเอาไว้ให้มากที่สุด เพื่อเป็นประโยชน์ต่อการนำไปศึกษาด้านภาษาศาสตร์ได้อีกทางหนึ่ง

6.2.3 ออนโทโลยีตำรายาโบราณอีสาน

ออนโทโลยีตำรายาโบราณอีสานเป็นองค์ประกอบหนึ่งของตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน พบว่า ช่วยให้ระดับความถูกต้องของคำอ่านภาษาไทยถิ่นอีสานเพิ่มสูงขึ้น เนื่องจากการจัดเก็บข้อมูลในออนโทโลยีมีการออกแบบและพัฒนาตามโครงสร้างการเขียนตำรายาโบราณอีสาน ทำให้คำศัพท์ถูกนำมาจัดกลุ่มอยู่ในบริบทเดียวกัน เช่น ชื่อยา รสยา อาการป่วย/ชื่อโรค เกษัชวัตถุ วิธีเตรียมยา/ปรุงยา ปริมาณ หน่วยวัดและชั่งตวง วิธีใช้ยา คำรักษา เป็นต้น และการสร้างความสัมพันธ์ระหว่างคำศัพท์ในออนโทโลยีใช้ความสัมพันธ์แบบคำที่มักเกิดขึ้นด้วยกันในตำแหน่งด้านซ้ายและขวา (คำบริบทซ้ายและขวา) ช่วยให้สามารถเจาะจง หรือระบุคำอ่านภาษาไทยถิ่นอีสานที่มีความถูกต้องสูงขึ้น

การใช้จัดเก็บคำศัพท์และคำบริบทในออนโทโลยี พบว่า การสอบถามข้อมูลใช้เวลาประมวลผลค่อนข้างนาน หากจำนวนคำศัพท์และความสัมพันธ์ระหว่างคำศัพท์มีจำนวนมาก เวลาประมวลผลจะยิ่งนานมากขึ้น นอกจากนี้พบว่า ความผิดพลาดในการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานยังมีลักษณะคล้ายคลึงกับการใช้คลังศัพท์ คือ กรณีเป็นคำที่ไม่รู้จัก จะมีสาเหตุมาจากไม่มีคำอ่านภาษาไทยถิ่นอีสานที่สอบถามในออนโทโลยี หรือมีคำอ่านแต่คำบริบทไม่ตรงกับออนโทโลยี และกรณีแปลงคำอ่านผิดมีสาเหตุจากคำอ่านภาษาไทยถิ่นอีสานบางคำมีบริบทเหมือนกันแต่อยู่ต่างคลาสกัน (ความหมายต่างกัน) เช่น บาด (คำปริวรรต) มีคำอ่าน 3 แบบ และอยู่คนละคลาส ได้แก่ “บาด” (แผล) “บาด” (ครั้ง) “บาท” (มาตราชั่งตวง) โดยพบว่า “บาด” คำที่ 2 และ 3 มีคำบริบทเหมือนกัน เช่น “บาด๑” (1 ครั้ง) และ “บาท๑” (น้ำหนัก 1 บาท) ทำให้การใช้คำบริบทที่อยู่ติดกันด้านซ้าย หรือขวา ข้างละ 1 คำ ยังมีข้อผิดพลาด รวมถึงคำบริบทบางคำไม่สามารถระบุคำอ่านที่ถูกต้องได้ เนื่องจากเป็นคำที่เกิดได้กับทุกคำในทุกส่วนเนื้อหา นอกจากนี้ยังพบว่า ตำแหน่งคำบริบทที่ช่วยเจาะจงคำอ่านที่ถูกต้องนั้นจะถูกเขียนอยู่ในตำแหน่งที่ไกลออกไปมากกว่า 1 คำ กล่าวคือ บางคำอยู่ในตำแหน่งห่างออกไป 2 หรือ 3 คำ หรือไม่สามารถเจาะจงระยะห่างได้อย่างชัดเจน เช่น อาจจะอยู่ต้นประโยค หรือท้ายประโยค เป็นต้น ดังนั้นการทำคำอ่านที่ถูกต้องสำหรับคำกลุ่มนี้จะต้องดูจากตำแหน่งการปรากฏของคำอ่านนั้นว่า เขียนอยู่ในเนื้อหาหมวดใดในตำรายา เช่น ส่วนของชื่อโรคและอาการของโรค ส่วนของรายชื่อสมุนไพรที่ใช้รักษา ส่วนของวิธีเตรียมยา/ปรุงยา หรือส่วนของวิธีใช้ยา จึงจะสามารถทำคำอ่านถูกต้อง เป็นต้น

6.3 ข้อจำกัดของการวิจัย

การประเมินประสิทธิภาพความถูกต้องการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน มีข้อจำกัดในการวิจัยคือ

6.3.1 ชุดข้อมูลมีปริมาณค่อนข้างจำกัด เนื่องจากข้อมูลสำหรับนำมาใช้เป็นข้อมูลชุดฝึกและข้อมูลชุดทดสอบประสิทธิภาพของกรอบการทำงานและขั้นตอนวิธีการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานที่นำเสนอในงานวิจัยนี้ (คำรับยาโบราณอีสานแบบคำปริวรรต) จะต้องมีภาพถ่ายเอกสารโบราณคำรายต้นฉบับและจะต้องสามารถติดต่อกับผู้ปริวรรตได้เท่านั้น เพื่อให้สามารถตรวจสอบความถูกต้องของข้อมูลได้ ทำให้ชุดข้อมูลที่สามารถนำมาใช้ได้มีปริมาณค่อนข้างจำกัด แม้ว่าจะมีคำรายาโบราณอีสานที่บันทึกด้วยอักษรไทยน้อยอีกเป็นจำนวนมาก แต่การปริวรรตจะต้องใช้ทรัพยากรมาก ทั้งผู้เชี่ยวชาญ เวลา และงบประมาณ งานวิจัยนี้จึงเลือกใช้ข้อมูลจากคำรายาโบราณอีสานที่ศูนย์อนุรักษ์ศิลปะและวัฒนธรรมอีสาน มหาวิทยาลัยมหาสารคาม ที่ได้ปริวรรตและพิมพ์เผยแพร่ไว้เท่านั้น เนื่องจากเป็นชุดข้อมูลที่มีภาพถ่ายเอกสารโบราณคำรายต้นฉบับและสามารถติดต่อกับผู้ปริวรรตได้ ชุดข้อมูลที่รวบรวมได้จึงมีปริมาณค่อนข้างจำกัด

6.3.2 การสร้างข้อมูลชุดทดสอบ คำปริวรรตแต่ละประเภทไม่สามารถสร้างให้มีจำนวนเท่ากันได้ เนื่องจากคำรายาโบราณอีสานที่ปริวรรตมาจากคำรายาอักษรไทยน้อยที่มีภาพคำรายต้นฉบับ มีจำนวนค่อนข้างน้อย และคำปริวรรตบางคำโดยทั่วไปมักจะพบการใช้บ่อยอยู่แล้ว

6.3.2 การอธิบายความหมายของคำศัพท์ในคลังศัพท์ มีบางคำที่ไม่สามารถให้คำอธิบายและชนิดของคำได้ เนื่องจากไม่มีผู้ที่ทราบคำอ่านและคำแปล

6.4 การประยุกต์ใช้ผลการวิจัย

กรอบการทำงานที่นำเสนอในงานวิจัยนี้ สามารถนำไปประยุกต์ใช้สำหรับการพัฒนาระบบแปลงคำปริวรรตของอักษรโบราณ หรือการแปลภาษาด้วยเครื่อง ทั้งในอักษรไทยโบราณ และอักษรโบราณของชาติอื่น ๆ ที่มีลักษณะภาษาต้นฉบับและภาษาเป้าหมายเป็นภาษาเดียวกัน แต่อักษรที่บันทึกในเอกสารต้นฉบับมีจำนวนน้อยกว่าอักษรที่ใช้บันทึกภาษาเป้าหมาย เช่น อักษรธรรมอีสานที่บันทึกเป็นภาษาไทยถิ่นอีสาน อักษรธรรมล้านนาที่บันทึกเป็นภาษาไทยถิ่นเหนือ เป็นต้น โดยการเปลี่ยนคำปริวรรต คำอ่านในคลังศัพท์ และคำศัพท์ในออนโทโลยี ให้ตรงกับอักษรและภาษาที่ต้องการนำไปประยุกต์ใช้

6.5 ข้อเสนอแนะในการวิจัยครั้งต่อไป

6.5.1 การเพิ่มประสิทธิภาพของคลังศัพท์ คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐาน มีข้อเสนอแนะดังนี้

เนื่องจากคลังคำในงานวิจัยนี้ รวบรวมคำศัพท์จากตำราयाโบราณอีสานที่ปริวรรตมาจากตำราयाโบราณอีสานที่บันทึกในใบลานด้วยอักษรไทยน้อยเท่านั้น ทำให้ได้คำศัพท์ค่อนข้างน้อย และส่งผลให้การตัดคำและการทำคำอ่านเกิดความผิดพลาดสูง ดังนั้น หากเพิ่มจำนวนคำศัพท์ให้มากขึ้นโดยการรวบรวมคำศัพท์จากตำราयाโบราณอีสานที่ปริวรรตมาจากอักษรธรรมอีสาน ที่พบว่า เป็นอักษรที่ใช้บันทึกตำราयाโบราณอีสานเช่นเดียวกับอักษรไทยน้อย ซึ่งนอกจากจะช่วยเพิ่มจำนวนคำศัพท์แล้ว ยังช่วยให้คลังศัพท์นี้สามารถนำไปใช้แปลงคำปริวรรตของตำราयाโบราณอีสานที่บันทึกด้วยอักษรธรรมอีสานได้อีกด้วย

6.5.2 ออนโทโลยีตำราयाโบราณอีสาน

เนื่องจากการใช้ออนโทโลยีตำราयाโบราณอีสานในการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสานยังมีความผิดพลาด และการประมวลผลใช้เวลาค่อนข้างนาน จึงมีข้อเสนอแนะในการทำวิจัยต่อไป แบ่งเป็น 2 แนวทาง ได้แก่

6.5.2.1 การลดความผิดพลาดของการแปลงคำปริวรรตเป็นคำอ่านภาษาไทยถิ่นอีสาน ในกรณีเป็นคำที่ไม่รู้จักที่มีสาเหตุจากบริบทของคำอ่านที่สอบถามไม่ตรงกับข้อมูลในออนโทโลยี โดยใช้วิธีเพิ่มเงื่อนไขการวิเคราะห์คำอ่านและคำบริบทที่สอบถามว่า อยู่ในคลาสเดียวกันหรือไม่ เนื่องจากคำอ่านที่ถูกต้องนอกจากจะสอดคล้องกับคำบริบทที่อยู่รอบข้างแล้ว ตำแหน่งที่อยู่ของคำอ่าน (อยู่ในคลาสเดียวกัน) รวมถึงการใช้ POS มาช่วยวิเคราะห์คำอ่านที่เหมาะสม จะสามารถช่วยให้การเลือกคำอ่านมีความถูกต้องเพิ่มขึ้น

6.5.2.2 การลดเวลาในการประมวลผล จากการจัดเก็บคำศัพท์และความสัมพันธ์กับคำบริบทไว้ในโครงสร้างไฟล์แบบ RDF ในงานวิจัยนี้ พบว่า ยิ่งจำนวนคำศัพท์และความสัมพันธ์มีจำนวนมากและซับซ้อน ยิ่งเพิ่มเวลาในการประมวลผลข้อมูลมากขึ้น ดังนั้น หากต้องการเพิ่มประสิทธิภาพด้านความเร็วในการประมวลผล ควรจัดเก็บคำศัพท์และความสัมพันธ์ระหว่างคำบริบทไว้ในโครงสร้างข้อมูลอื่นที่รองรับการจัดเก็บคำศัพท์และคำบริบทตามแนวคิดแบบออนโทโลยี

6.5.3 คลังข้อความคู่ขนาน คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐานขนาดใหญ่

ในปัจจุบันการประมวลผลภาษาธรรมชาติด้วยเทคนิควิธีสร้างตัวแบบจากการเรียนรู้ด้วยเครื่องเป็นเทคนิคที่ได้รับการยอมรับว่ามีประสิทธิภาพสูง แต่เนื่องจากเป็นเทคนิควิธีที่ต้องอาศัยคลังข้อความคู่ขนานขนาดใหญ่มากสำหรับใช้เป็นฐานความรู้ และสร้างตัวแบบสำหรับปริวรรตและแปลเอกสารโบราณแบบอัตโนมัติด้วยเครื่อง ในขณะที่ปัจจุบันยังไม่มีคลังข้อความคู่ขนานระหว่างอักษรและภาษาโบราณ-อักษรและภาษาปัจจุบันให้ใช้ ดังนั้น การพัฒนาคลังข้อความคู่ขนาน คำปริวรรต-คำอ่านภาษาไทยถิ่นอีสาน-คำอ่านภาษาไทยมาตรฐานจะช่วยสนับสนุนให้เกิดการพัฒนาเทคนิควิธีใหม่ ๆ ในการปริวรรตเอกสารโบราณที่บันทึกด้วยอักษรไทยน้อยแบบอัตโนมัติด้วยเครื่อง

6.5.4 ระบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน

จากการพัฒนาระบบต้นแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสาน พบว่า ระบบต้นแบบช่วยอำนวยความสะดวกในการปริวรรตตำรายาโบราณอีสานเป็นอย่างมากทั้งในกลุ่มผู้เชี่ยวชาญและผู้ใช้ทั่วไป โดยในกลุ่มผู้เชี่ยวชาญ พบว่า ช่วยให้การปริวรรตตำรายาโบราณอีสานมีความสะดวกยิ่งขึ้น ลดขั้นตอนการทำงาน สามารถบันทึกจัดเก็บข้อมูลไว้อย่างเป็นระบบ และการปรับปรุงแก้ไขได้ง่าย สำหรับกลุ่มผู้ใช้ทั่วไป ระบบต้นแบบช่วยสนับสนุนการศึกษาค้นคว้าตำรายาโบราณอีสานให้มีความสะดวกรวดเร็วขึ้น โดยมีข้อเสนอแนะในการพัฒนาระบบเพิ่มเติม ได้แก่ ข้อมูลนำเข้าควรใช้ภาพถ่ายใบลานตำรายาได้ ระบบสามารถแปลงจากภาพอักษรโบราณเป็นอักษรโบราณ (OCR) ที่แก้ไขด้วยคอมพิวเตอร์ได้ สามารถปริวรรตหรือถอดจากอักษรโบราณเป็นคำปริวรรตอักษรไทยปัจจุบันได้โดยอัตโนมัติ คำปริวรรตสามารถแสดงเครื่องหมายที่อธิบายวิธีประสมอักษรในเอกสารต้นฉบับได้ เช่น เครื่องหมายขีดบน แสดงถึง ถอดมาจากพยัญชนะประสม เช่น หิมาก เครื่องหมายจุดใต้พยัญชนะ หมายถึง ถอดมาจากพยัญชนะควบกล้ำที่เขียนอยู่ใต้พยัญชนะต้น เช่น กุลาว เป็นต้น

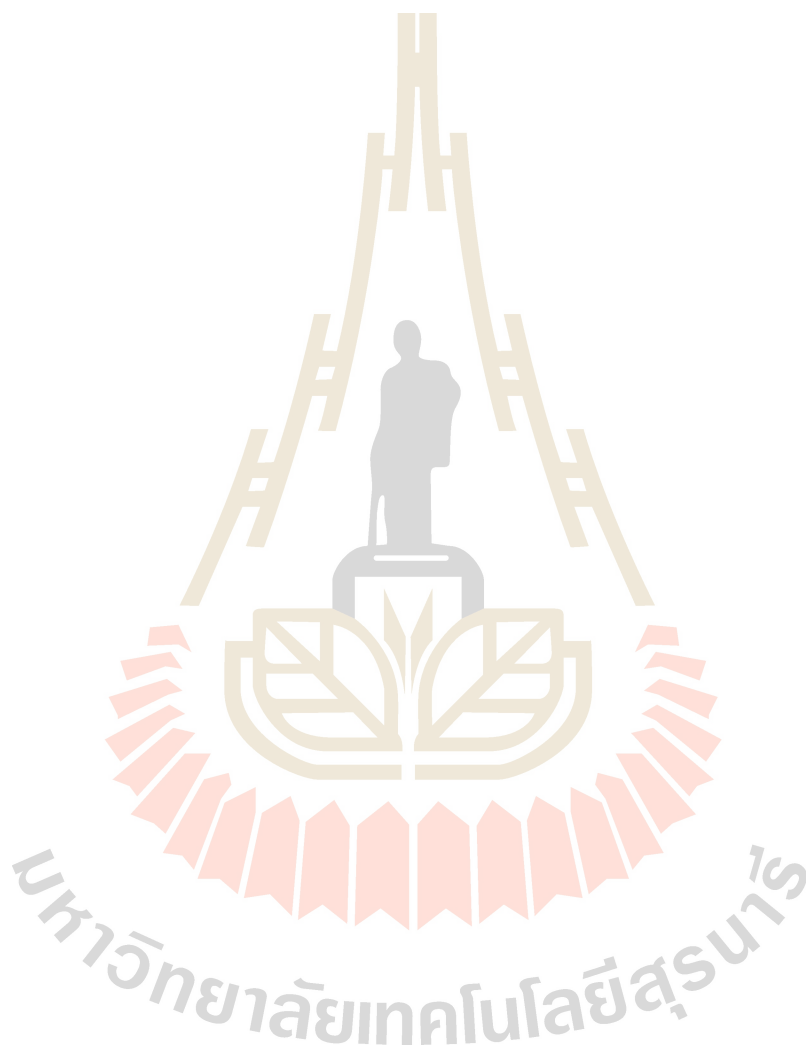
6.6 องค์ความรู้ใหม่ในงานวิจัย (Contribution)

งานวิจัยเรื่องการพัฒนากรอบการทำงานการแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาไทยถิ่นอีสานในครั้งนี้ มีข้อค้นพบใหม่ทางวิชาการ 3 ประการ ดังนี้

6.6.1 กรอบการทำงานและขั้นตอนวิธีในการแปลงคำปริวรรตของอักษรไทยน้อย (อักษรโบราณประเภทหนึ่ง) เป็นคำอ่านภาษาไทยถิ่นอีสาน สำหรับการพัฒนากระบวนการปริวรรตเอกสารโบราณอักษรไทยน้อยด้วยเครื่อง

6.6.2 คลังศัพท์ภาษาไทยถิ่นอีสาน หมวดคำราชาโบราณอีสาน ที่มีคำศัพท์หลักเป็นคำปริวรรต สำหรับใช้ในการตัดคำและทำคำอ่านภาษาไทยถิ่นอีสาน การสะกดคำตามพจนานุกรมสำหรับระบบแปลภาษาด้วยเครื่อง ประเภทเอกสารโบราณอีสาน

6.6.3 ออนโทโลยีคำศัพท์ภาษาไทยถิ่นอีสาน หมวดคำราชาโบราณอีสาน ซึ่งเป็นฐานความรู้ทางภาษาคู่ขนานของคลังศัพท์ ที่ช่วยเพิ่มประสิทธิภาพความถูกต้องในการทำคำอ่านให้สูงขึ้น



รายการอ้างอิง

- กมล แก้วปลั่ง. (2556). อักษรไทยน้อย ภาษาผูกพันวัฒนธรรมไทย-ลาว. วารสารวัฒนธรรม. 52 (4).
- กรรณิการ์ วิมลเกษม. (2548). หลักการปริวรรตอักษรโบราณที่พัฒนามาจากอักษรของอินเดีย. วารสารดำรงวิชาการ. 4(1): 120-135.
- กองยุทธศาสตร์และแผนงาน. (2561). แผนยุทธศาสตร์ชาติระยะ 20 ปี ด้านสาธารณสุข (พ.ศ. 2560-2579) ฉบับปรับปรุงครั้งที่ 2. นนทบุรี: สำนักปลัดกระทรวงสาธารณสุข กระทรวงสาธารณสุข.
- กำชัย ทองหล่อ. (2543). หลักภาษาไทย. กรุงเทพฯ: รวมสาส์น.
- กิตติวัฒน์ สัตนาโค. (2542). อักษรวิเศษในอักษรอีสาน-ล้านนา กับศิลาจารึก หลักที่ 1. วารสารมหาสารคาม, 17 (2): 64 - 70.
- งามนิจ อัจฉินทร์. (2557). เทคโนโลยีเว็บแบบสื่อความหมาย (พิมพ์ครั้งที่ 3). ขอนแก่น: คลังนานาวิทยา.
- จันทร์แดง คำลือหาญ. (2535). อักษรวิเศษในอักษรธรรมและไทยน้อย. วารสารมหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม, 11 (1): 33 - 38.
- ณัฐิมา สุระเดช. (2554). การวิเคราะห์โครงสร้างคำของภาษาล้านนาโดยใช้เทคนิคการทำเหมืองกราฟ. วิทยาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่.
- ทัศนวรรณ ศูนย์กลาง สุนีย์ พงษ์พินิจภิญโญ และวิภาวดี ม่วงอัน. (2558). การแปลภาษาไทย-อีสาน โดยใช้ฐานกฎ. วารสารวิชาการคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏรำปาง, 8 (2) : 23-33.
- ชนานันท์ ตรงดี. (2554). รายงานวิจัยเรื่อง ไวยากรณ์ไทยฉบับครอบคลุมภาษาย่อย. กรุงเทพฯ : สำนักงานกองทุนสนับสนุนการวิจัย.
- ธวัช ปุณโณทก. (2549). อักษรไทยโบราณ ลายสือไทย และวิวัฒนาการของอักษรของชนชาติไทย. กรุงเทพฯ: สำนักห้องสมุดแห่งจุฬาลงกรณ์มหาวิทยาลัย.
- นงคันุช ไพบูลย์. (2552). พจนานุกรมอักษรธรรมอีสาน. การศึกษาอิสระ ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.
- นิกร ยาทรม. (2554). รายงานการวิจัยเรื่อง พจนานุกรมอิเล็กทรอนิกส์ ล้านนา-ไทย-อังกฤษ. สำนักงานการวิจัยแห่งชาติ.
- นิสาชล จำนงศรี. (2562). การพัฒนาออนโทโลยีด้านยาสมุนไพรที่บันทึกไว้ในเอกสารโบราณ. วารสารสารสนเทศศาสตร์, 37 (4): 17 - 43.

- บุญเกิด พิมพ์วรรณเมธากุล และนภาพร พิมพ์วรรณเมธากุล. (2545). **พจนานุกรมภาษาถิ่นอีสาน: เว้าอีสาน**. ขอนแก่น : คลังนาวิทยา.
- ประเสริฐ ฒ นกร. (2541). **สารนิพนธ์ ประเสริฐ ฒ นกร**. กรุงเทพฯ: พิมพ์ลักษณ์.
- ปรีชา พิณทอง. (2532). **สารานุกรม ภาษาอีสาน - ไทย - อังกฤษ**. กรุงเทพฯ : โรงพิมพ์ศิริธรรม.
- วีระ สุขสังข์. (2548). **พจนานุกรมภาษาอีสาน ฉบับคำพูน บุญทวี**. กรุงเทพฯ : โป๊ยเซียน.
- ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (ม.ป.ป.). **พจนานุกรมอิเล็กทรอนิกส์ไทย-อังกฤษเล็กจิตรอน เวอร์ชัน 3.0**. จาก https://lexitron.nectec.or.th/2009_1/index.php?q=lookup/form/submit
- อุปกิตศิลปสาร (นิ่ม กาญจนาชีวะ), พระยา. (2544). **หลักภาษาไทย (อักขรวิธี วจีภาค วายก สัมพันธ์ ฉันทลักษณ์)**. กรุงเทพฯ: ไทยวัฒนาพานิช.
- พิเชฐ ตันติพัฒน์ภิญโญ. (2551). **การพัฒนาโปรแกรมพจนานุกรมอิเล็กทรอนิกส์ล้านนา-ไทย**. การค้นคว้าแบบอิสระ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่. เชียงใหม่: บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่
- พิชญ์นาฏ สุทธิสมณ์. (2557). รายงานคำศัพท์พื้นฐาน 1,000 คำ ในการเรียนการสอนรายวิชา วิทยาศาสตร์และเทคโนโลยีเพื่อคุณภาพชีวิตและสังคม (SC 103) สำหรับนักศึกษาชาวจีนที่ใช้ภาษาไทยเป็นภาษาที่สองของมหาวิทยาลัยธุรกิจบัณฑิตย์. **วารสารสุทธิปริทัศน์**, 29 (92): 294 - 303.
- พิมพ์พรรณ ไพบูลหวั่งเจริญ. (2542). การปริวรรตเอกสารโบราณ. **วารสารมนุษยศาสตร์ปริทรรศน์ มหาวิทยาลัยศรีนครินทรวิโรฒ ปีที่ 21**: 25 - 30.
- พิมพ์พรรณ ไพบูลหวั่งเจริญ. (2543). **ใบจุ่ม: สารนิเทศบนสิ่งทอ**. กรุงเทพฯ: สำนักหอสมุดแห่งชาติ กรมศิลปากร.
- มาลี กาบมาลา ลำปาง แม่นมดัย และ ครรชิต มาลัยวงศ์. (2549). **ออนโทโลยี: แนวคิดการพัฒนา**. **วารสารบรรณารักษศาสตร์และสารนิเทศศาสตร์ มหาวิทยาลัยขอนแก่น ปีที่ 24 (1-3 มกราคม-ธันวาคม 2549)**: 24 - 47.
- รัตนา จันทร์เทาว์. (2558). **พจนานุกรม 3 ภาษา (ไทน้อย-ลาว-ไทย)**. มหาวิทยาลัยขอนแก่น.
- วิโรจน์ อรุณมานะกุล. (2553). **ภาษาศาสตร์คลังข้อมูล (พิมพ์ครั้งที่ 2)**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- วิไลศักดิ์ กิ่งคำ. (2556). **ภาษาไทยถิ่น**. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- ศิริลักษณ์ โรจกิจอำนวย. (2552). **การออกแบบและบริหารฐานข้อมูล (พิมพ์ครั้งที่ 3)**. กรุงเทพฯ: ดวงกมลสมัย.

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (2555). **คู่มือการใช้งาน Hozo-Ontology**

Editor.

จาก https://lst.nectec.or.th/oam_en/document_doc/Hozo_ThaiManual_25550123.pdf

สมเด็จพระมหาวีรวงศ์ (ติสฺสมหาเถร). (2541). **พจนานุกรมภาคอีสาน-ภาคกลาง ฉบับปณิธาน สมเด็จพระมหาวีรวงศ์ (ติสฺสมหาเถร).** กรุงเทพฯ: สำนักพระราชเลขธิการ.

สมชาย ปราการเจริญ. (2557). **ออนโทโลยี: ทางเลือกของการพัฒนาฐานความรู้ในรูปแบบเชิงเนื้อหา.** กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, คณะวิทยาศาสตร์ ประยุกต์.

สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2564). **(ร่าง)แผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 13.** กรุงเทพฯ: สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ.

สำลี รักสุทธี. (2554). **พจนานุกรมภาษาไทยกลาง.** กรุงเทพฯ: พัฒนาศึกษา.

สุภณ สมจิตศรีปัญญา. (2525). **อักษรไทยน้อย อักษรกาบ อักษรสร้อย.** (อนุสรณ์งานพระราชทานเพลิงศพ พระราชศีลโสภิต (ทา พุทฺธสรมหาเถร) เจ้าคณะจังหวัด มหาสารคาม 20 มิถุนายน 2525.

โสภณ พรหมโสดา. (2553). **การแปลอักษรธรรมอีสานให้เป็นภาษาไทย.** การศึกษาอิสระ ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น.

หอสมุดแห่งชาติ. (2545). **ประมวลคำศัพท์ภาษาลาวในเอกสารโบราณ เล่ม 1-5.** กรุงเทพฯ: ประชาชน จำกัด.

อมรา ประสิทธิ์รัฐสินธุ์. (2544). **ทฤษฎีไวยากรณ์.** กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

อักษรไทยน้อย. **วารสารศิลปากร กรมศิลปากร, ฉบับที่ 1 มกราคม - กุมภาพันธ์ 2558: 45 - 59.**

อังคณา ทองพูน พัฒนสร. (2560). **การพัฒนารายการคำศัพท์วิชาการทางภาษาอังกฤษธุรกิจ โดยใช้วิธีการด้านภาษาศาสตร์คลังข้อมูล.** *วารสารมนุษยศาสตร์ สังคมศาสตร์*, 34 (2): 1 - 31.

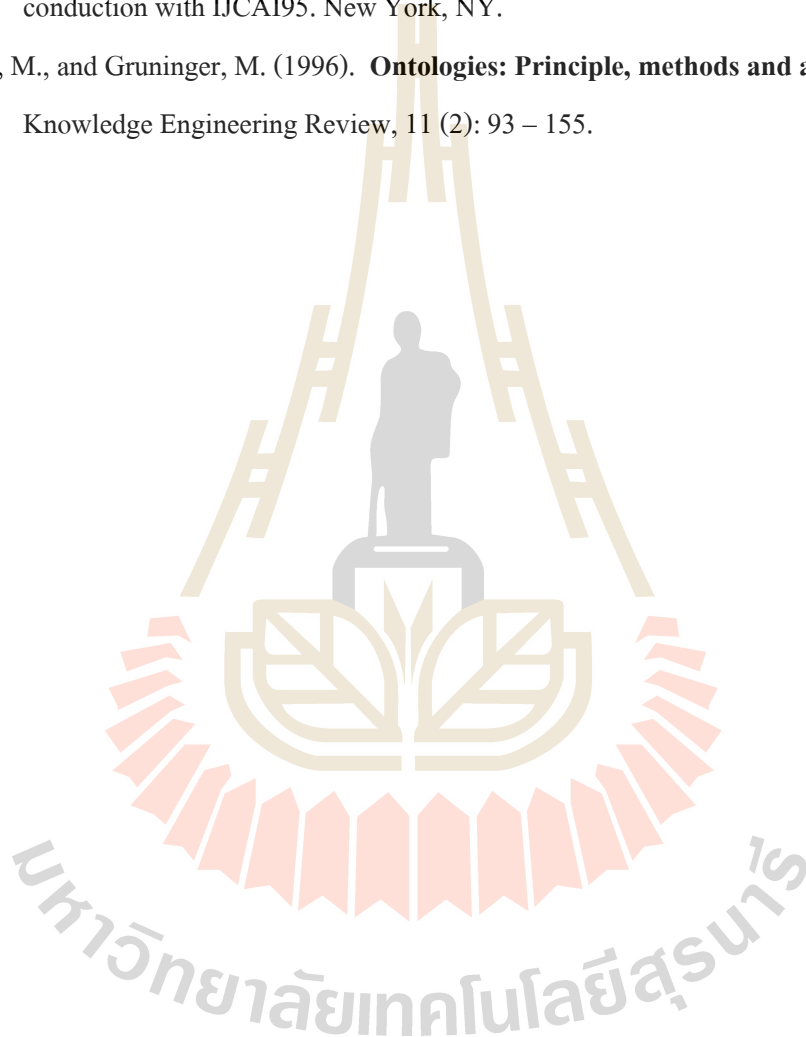
อิสรภาพ ล้อรัตนไชยรงค์. (2560). **การสร้างคลังศัพท์บอกความรู้สึกในภาษาไทยจากบทวิจารณ์.** อักษรศาสตรมหาบัณฑิต สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

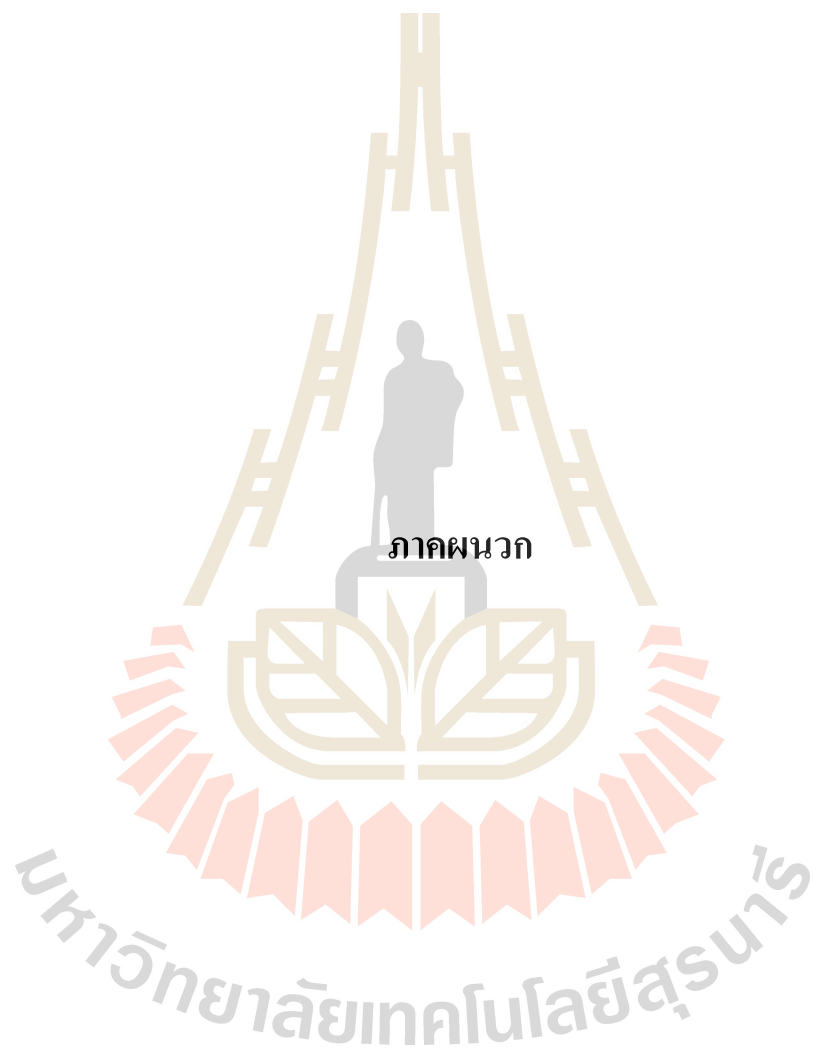
Abdelaal, N. M., and Alazzawie, A. (2020). **Machine Translation: The Case of Arabic-English Translation of News Texts.** *Theory and Practice in Language Studies*, 10 (4): 408 - 418.

- Azizan, A., Jamal, N. N. S. A., Abdullah, M. N., Mohamad, M., and Khairudin, N. (2019). **Lexicon-Based Sentiment Analysis for Movie Review Tweets**. In Conference, 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2019 (pp 132 - 136). Ipoh, Malaysia.
- Bhadwal N., Agrawal P., and Madaan V. (2020). **Bilingual Machine Translation System Between Hindi and Sanskrit Languages**. Scalable Computing: Practice and Experience, 21 (3): 543 – 553.
- Brank, J., Grobelnik, M., and Mladenic, D. (2005). **A survey of ontology evaluation techniques**. In Proceedings of the Conference on Data Mining and Data Warehouses SiKDD 2005. From https://www.researchgate.net/publication/228857266_A_survey_of_ontology_evaluation_techniques
- Control Device Division Enegate Co, Ltd. (2011). **Hozo – Ontology Editor (Distributed Development Version) Operating Manual [On-line]**. From http://www.hozo.jp/HozoManual_en_20140317.pdf
- George, A. M. (1995). WordNet: A Lexical Database for English. **Communications of the ACM**, 38 (11) : 39 - 41
- Ghosh, S., and Gunning, D., (2019). **Natural Language Processing Fundamentals**. UK: Packt Publishing Ltd.
- Hirst, G. (2009). Ontology and the Lexicon. In **Staab S., Studer R. (eds) Handbook on Ontologies** International Handbooks on Information Systems. Springer, Berlin, Heidelberg. From https://doi.org/10.1007/978-3-540-92673-3_12
- Intasorn, J., Gertphol, S., and Sammapun ,U. (2021). **Thai Sentiment Lexicon Construction**. In Conference 2021 13th International Conference on Knowledge and Smart Technology (KST), 2021 (pp 123 - 128). Chonburi, Thailand.
- Kituku, B., Muchemi, L., and Nganga,W. (2016). A Review on Machine Translation Approaches. **Indonesian Journal of Electrical Engineering and Computer Science**, 1 (1): 182- 190.
- Kolle, P., Bhagat, S., Zade, S., Dand, B., and Lifna, C. S. (2018). **Ontology Based Domain Dictionary**. In 2018 International Conference on Smart City and Emerging Technology (ICSCET), IEEE (pp 1–4).

- Lakkhanawannakun, P., and Seresangtakul, P. (2014a). **Improved Isarn Dharma alphabets to Thai language translation using longest syllable matching with named entities recognition**. In 5th International Conference on Advanced Computer Control.
- Lakkhawannakun P., and Seresangtakul P. (2014b). **Isarn Dharma Alphabets to Thai Language Translation by ATNs**. In: Jeong H., S. Obaidat M., Yen N., Park J. (eds) Advances in Computer Science and its Applications. Lecture Notes in Electrical Engineering: 775-782.
- Noy, N.F., and McGuinness, D.L. (2001). **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 (pp 1 - 25).
- Okumura, A., and Hovy, E. (1994). **Building Japanese-English Dictionary based on Ontology for Machine Translation**. In Proceedings of the workshop on Human Language Technology March 1994 (pp 141 – 146).
From <https://doi.org/10.3115/1075812.1075841>
- Phaiboon, N., and Seresangtakul, (2017). **Isarn Dharma Alphabets lexicon for natural language processing**. In 2017 9th International Conference on Knowledge and Smart Technology (KST), 2017 (pp 211 - 215), doi: 10.1109/KST.2017.7886108.
- Princeton University. (2010). **About WordNet** [On-line]. From <https://wordnet.princeton.edu/>.
- Somsap, S., and Seresangtakul, P. (2020). **Isarn Dharma Word Segmentation Using a Statistical Approach with Named Entity Recognition**. In: Jeong H., S. Obaidat M., Yen N., Park J. (eds) Advances in Computer Science and its Applications. Lecture Notes in Electrical Engineering, vol 279. Springer, Berlin, Heidelberg.
- Sornlertlamvanich, V. (1995). **Papers on Natural Language Processing: Multi-lingual Machine Translation and Related Topics (1987-1994)**. National Electronics and Computer Technology Center, National Science and Technology Development Agency Ministry of Science.
- Sornlertlamvanich, V. (2008). **From LEXiTRON to Asian WordNet on Collaborative Development Platform, 10 Years LEXiTRON**, NECTEC-ACE, Bangkok, September 2008.

- Thoongsup, S., Robkop, K., Mokarat, C., Sinthurahat, T., Charoenporn, T., Sornlertlamvanich, V., and H. Isahara. (2009). **Thai WordNet Construction**. In Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009 (pp 139 – 144).
- Uschold, M. & King, M. (1995). **Towards a methodology for building ontologies**. Paper presented at the Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI95. New York, NY.
- Uschold, M., and Gruninger, M. (1996). **Ontologies: Principle, methods and applications**. Knowledge Engineering Review, 11 (2): 93 – 155.







ภาคผนวก ก

แบบฟอร์มบันทึกการสัมภาษณ์

มหาวิทยาลัยเทคโนโลยีสุรนารี

แบบฟอร์มบันทึกการสัมภาษณ์

เรื่อง ปัญหาการปริวรรตอักษรโบราณและการปริวรรตคำราชาอิสานโบราณ

ที่บันทึกในใบลานด้วยอักษรไทยน้อย

วัตถุประสงค์ในการสัมภาษณ์

การสัมภาษณ์นี้มีวัตถุประสงค์เพื่อศึกษาปัญหาการปริวรรตอักษรโบราณ และปัญหาการปริวรรตคำราชาอิสานโบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย เป็นส่วนหนึ่งของคุณฉวีนิพนธ์ เรื่อง การใช้ออนโทโลยีเป็นฐานในการแปลงคำปริวรรตอักษรไทยน้อยเป็นคำอ่านภาษาอิสาน (An Ontology-based Approach for Transliterating Thainoi Characters into Isan Language Pronunciation) ข้อมูลที่ได้จากการสัมภาษณ์นี้จะประโยชน์อย่างยิ่งต่อการออกแบบและพัฒนาคลังคำคำราชาอิสานโบราณ การพัฒนาออนโทโลยีคำราชาอิสานโบราณ และการออกแบบและพัฒนาตัวแบบแปลงคำปริวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาอิสาน รวมถึงการพัฒนาเครื่องมือในการปริวรรตคำราชาอิสานโบราณที่บันทึกด้วยอักษรไทยน้อยเป็นคำอ่านภาษาอิสานแบบอัตโนมัติต่อไป

ผู้วิจัยขอขอบพระคุณในความอนุเคราะห์ของผู้ให้สัมภาษณ์ทุกท่าน ที่เสียสละเวลาให้สัมภาษณ์และเล่าประสบการณ์ด้านการปริวรรตอักษรโบราณ มา ณ โอกาสนี้

ด้วยความเคารพเป็นอย่างสูง

(นางสาวจินตนา พลศรี)

นักวิจัย

คำชี้แจง แบบสอบเป็นแบบกึ่งโครงสร้าง คำถามแบ่งเป็น 4 ตอน รวบรวมข้อมูลด้วยวิธีสัมภาษณ์
และบันทึกข้อมูลโดยผู้วิจัย

ตอนที่ 1. ข้อมูลพื้นฐานผู้ให้สัมภาษณ์

1. รหัสผู้ให้สัมภาษณ์.....
2. หน่วยงาน.....
3. ตำแหน่ง.....
4. ระยะเวลาที่ทำงานในตำแหน่งนักภาษาโบราณ..... ปี..... เดือน.....
5. อายุ.....ปี
6. ระดับการศึกษา สาขาวิชาที่จบ.....
ระดับปริญญา.....
7. สาขาวิชาที่เชี่ยวชาญ (เลือกได้มากกว่า 1 ข้อ)
 - อักษรวิทยาและภาษาโบราณ
 - ประวัติศาสตร์
 - ประวัติศาสตร์ศิลปะ
 - ภาษาศาสตร์
 - ภาษาและวรรณกรรมท้องถิ่น
 - อื่นๆ.....

ตอนที่ 2. ประสิทธิภาพและความสามารถในการปริวรรตเอกสารโบราณ

1. มีประสิทธิภาพด้านการปริวรรตเอกสารโบราณกี่ปี.....
2. จุดมุ่งหมายของการปริวรรตเอกสารโบราณคืออะไร.....
.....
.....

3. เอกสารโบราณที่เคยปริวรรตประเภทใดบ้าง มีประสบการณ์ในการปริวรรตเอกสารแต่ละประเภทกี่ปี?

- จารึก ประสบการณ์ ปี
- สมุดไทย ประสบการณ์ ปี
- คัมภีร์ใบลาน ประสบการณ์ ปี
- อื่น ๆ

4. อักษรโบราณที่รู้จักหรืออ่านได้ ได้แก่ อักษรใดบ้าง มีความเชี่ยวชาญอยู่ในระดับใด

ชื่ออักษรโบราณ	เชี่ยวชาญ	อ่านได้	รู้จักแต่อ่านไม่ได้
<input type="checkbox"/> อักษรไทย (ไทยสุโขทัย ไทยอยุธยา ไทยรัตนโกสินทร์)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรขอมไทย	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรปัลลวะ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรธรรมล้านนา	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรธรรมล้านช้าง	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรไทยน้อย	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรธรรมอีสาน	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรฝักขาม	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรเทวนาครี	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรมอญโบราณ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อักษรขอมโบราณ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อื่นๆ			



5. ภาษาโบราณที่รู้และเข้าใจความหมาย ได้แก่ภาษาใดบ้าง

ชื่อภาษาโบราณ	เชี่ยวชาญ	อ่านได้	รู้จักแต่อ่าน
ไม่ได้			
<input type="checkbox"/> บาลี	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> ไทยโบราณ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อีสาน	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> ล้านนา	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> อื่น ๆ			

6. เนื้อหาในเอกสารโบราณที่เคยปริวรรตเป็นเนื้อหาหมวดใด

- | | |
|---------------------------------------|---------------------------------|
| <input type="checkbox"/> วรรณกรรม | <input type="checkbox"/> กฎหมาย |
| <input type="checkbox"/> ศาสนา | <input type="checkbox"/> ตำรายา |
| <input type="checkbox"/> อื่น ๆ | |

ตอนที่ 3 ปัญหาการปริวรรตเอกสารโบราณ

1. ปัญหาด้านกายภาพของเอกสารโบราณ

1.1 ความไม่สมบูรณ์ของเอกสารโบราณ เช่น ชำรุด แตก หัก ลบเลือน เปราะเปื้อน เป็นปัญหาต่อการปริวรรตหรือไม่ อย่างไร

แนวทางแก้ไขปัญหาคือ



1.2 ปัญหาอื่นที่มีสาเหตุจากลักษณะกายภาพของเอกสารโบราณ คืออะไร แนวทางที่ใช้แก้ปัญหาคืออะไร

ปัญหา

แนวทางแก้ไขปัญหาคือ

ปัญหา

แนวทางแก้ไขปัญหาคือ

2. ปัญหาด้านรูปอักษรและภาษา

2.1 ลายมือเขียนที่มีความหลากหลายมีผลต่อปริวรรตหรือไม่ อย่างไร

2.2 รู้และเข้าใจภาษาถิ่นช่วยให้การปริวรรตง่ายขึ้นหรือไม่ อย่างไร

มีต่อ

2.3 คำศัพท์โบราณเป็นปัญหาต่อการปริวรรตหรือไม่ อย่างไร

2.4 ในเอกสารโบราณมีการใช้ คำพ้องรูป พ้องเสียง หรือคำที่สะกดต่างกันแต่มีความหมายเหมือนกัน หรือไม่ เป็นปัญหาต่อการปริวรรตหรือไม่ อย่างไร

2.5 เมื่อพบคำศัพท์ใหม่หรือไม่คุ้นเคยบ่อยขนาดไหน พจนานุกรมภาษาไทยและภาษาถิ่นสามารถช่วยได้หรือไม่ อย่างไร

2.6 ลีลาการเขียนตัวอักษรตัวเดียวกัน แต่รูปตัวอักษรไม่เหมือนกัน ทำให้เกิดปัญหาอย่างไร มีวิธีอย่างไรที่จะทำให้เดาว่าเป็นอักษรอะไร

2.7 พจนานุกรมภาษาถิ่นมีคำศัพท์เพียงพอและครอบคลุมคำศัพท์ในทุกทุกหมวดหรือไม่ ลักษณะพจนานุกรมสำหรับปริวรรตอักษรโบราณควรเป็นอย่างไร



2.8 ฐานข้อมูลคำศัพท์อักษรและภาษาโบราณหรือคลังคำศัพท์โบราณสำหรับใช้กับการ
ปริวรรตเอกสารโบราณ โดยเฉพาะมีหรือไม่ ลักษณะเป็นอย่างไร

2.9 การเขียนอักษรโบราณอื่นปะปนในเอกสารเดียวกัน เช่น อักษรธรรมอีสานเขียนปนกับ
อักษรไทยน้อย พบบ่อยหรือไม่ เป็นปัญหาต่อการปริวรรตหรือไม่ อย่างไร

2.10 อักษรโบราณที่ไม่มีวรรณยุกต์ เป็นปัญหาต่อการปริวรรตหรือไม่ การสะกดคำอ่านให้
ถูกต้องทำได้อย่างไร

2.11 การปริวรรตอักษรโบราณใช้เวลานานเกิดจากสาเหตุอะไร

2.12 การแก้ไขในเอกสารโบราณต้นฉบับในกรณีใดที่มักเป็นสาเหตุที่ทำให้ถ่ายทอดเป็น
อักษรไทยปัจจุบันผิด

2.13 การประสมอักษรเป็นคำแบบใดที่ทำให้ยากต่อการปริวรรต เพราะอะไร



2.14 การตัดคำมักที่มีการแยกพยัญชนะต้นและพยัญชนะตัวสะกดหรือตัวตามอยู่คนละบรรทัด คนละหน้า ทำให้เกิดปัญหาในการปริวรรตหรือไม่ อย่างไร

2.15 ความเชื่อวชาญในเนื้อหาที่บันทึกในเอกสารโบราณแต่ละหมวดไม่เท่ากัน เป็นปัญหาต่อการปริวรรตหรือไม่ อย่างไร

16. ปัญหาอื่น ๆ

3. ความรู้ ทักษะ และสิ่งจำเป็นสำหรับใช้ปริวรรตเอกสารโบราณ

3.1 ประสบการณ์ในการปริวรรตเอกสารโบราณเป็นระยะเวลานาน มีความสำคัญและจำเป็นต่อการการปริวรรตเอกสารโบราณหรือไม่ อย่างไร

3.2 ประสบการณ์ในปริวรรตเอกสารโบราณหลายหมวดหมู่ เช่น วรรณกรรม กฎหมาย หรือ ตำรายา เป็นต้น มีความจำเป็นหรือไม่ อย่างไร

3.3 การรู้จักและอ่านอักษรโบราณได้มากกว่า 1 ชนิด ช่วยให้การปริวรรตเอกสารโบราณ
ง่ายขึ้นหรือไม่ อย่างไร

3.4 การรู้จักและเข้าใจภาษาถิ่นเป็นอย่างดี ช่วยให้การปริวรรตเอกสารโบราณง่ายขึ้น
หรือไม่ อย่างไร

3.5 ทักษะด้านการใช้เทคโนโลยี เช่น การใช้โปรแกรมคอมพิวเตอร์ หรือ อินเทอร์เน็ต
ช่วยให้การปริวรรตทำได้ง่าย สะดวก รวดเร็ว หรือไม่ อย่างไร

3.6 อื่น ๆ

มหาวิทยาลัยเทคโนโลยีสุรนารี



**ตอนที่ 4 ปัญหาการปริวรรตอักษรไทยน้อยและตำรายาอีสานโบราณที่บันทึกในใบลานด้วย
อักษรไทยน้อย**

1. การปริวรรตอักษรไทยน้อยเป็นอักษรไทยปัจจุบัน มีปัญหาหรือลักษณะเฉพาะอะไรที่เหมือนหรือแตกต่างจากการปริวรรตอักษรโบราณอื่น หรือไม่ อย่างไร

.....

.....

.....

2. เมื่อเจอปัญหาการเกี่ยวกับการทำคำปริวรรต คำอ่าน ไม่ได้ หรือพบคำศัพท์ใหม่ที่ไม่รู้จักมาก่อน มีวิธีการแก้ปัญหาเหล่านั้นนี้อย่างไร

.....

.....

.....

3. ความยากในการปริวรรตอักษรไทยน้อยและการปริวรรตตำรายาอีสานโบราณที่บันทึกในใบลานด้วยอักษรไทยน้อย ได้แก่อะไรบ้าง มีสาเหตุจากอะไร

.....

.....

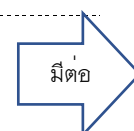
.....

4. จากข้อ 3. เมื่อเจอปัญหา/อุปสรรคดังกล่าว มีวิธีแก้ปัญหา อย่างไร

.....

.....

.....



5. ลักษณะฐานความรู้สำหรับใช้ปริวรรตอักษรไทยน้อยและปริวรรตคำราชาธิสาณโบราณที่บันทึกด้วยอักษรไทยน้อยควรมีลักษณะอย่างไร

.....

.....

.....

.....

.....

.....

6. ข้อคิดเห็นอื่น ๆ

.....

.....

.....

.....

.....

.....

วันที่ให้สัมภาษณ์ เวลา

มหาวิทยาลัยเทคโนโลยีสุรนารี



ภาคผนวก ข

แบบประเมินประสิทธิภาพการทำงานในการแปลงคำปริวรรต
ของอักษรไทยน้อยเป็นคำอ่านภาษาอีสาน

มหาวิทยาลัยเทคโนโลยีสุรนารี

แบบประเมินประสิทธิภาพการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อย

เป็นคำอ่านภาษาอีสาน

คำชี้แจง

1. แบบประเมินประสิทธิภาพการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อยนี้ เป็นส่วนหนึ่งของคุษฎีนิพนธ์เรื่อง การใช้ออนโทโลยีเป็นฐานในการแปลงคำปรัวรรตอักษรไทยน้อยเป็นคำอ่านภาษาอีสาน (An Ontology-based Approach for Transliterating Thainoi Characters into Isan Language Pronunciation) ของ นางสาวจินตนา พลศรี นักศึกษาหลักสูตรปรัชญาคุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

2. วัตถุประสงค์ของการประเมินเพื่อให้ทราบประสิทธิภาพของการทำงานที่พัฒนาขึ้น และใช้เป็นข้อเสนอแนะในการปรับปรุงการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาอีสานและการพัฒนาระบบปรัวรรตคำรายอีสาน โบราณที่บันทึกในใบลานด้วยอักษรไทยน้อยแบบออตโนมัตินี้ ให้มีประสิทธิภาพยิ่งขึ้น

3. การทำแบบประเมินจะทำหลังผู้ประเมินได้ทดลองใช้ระบบต้นแบบที่ผู้วิจัยนำมาให้ทดลองใช้ก่อน

4. เกณฑ์การให้คะแนนประสิทธิภาพแบ่งเป็น 5 ระดับ ได้แก่ ดีมาก (5 คะแนน) ดี (4 คะแนน) พอใช้ (3 คะแนน) น้อย (2 คะแนน) ควรปรับปรุง (1 คะแนน)

5. การประเมินจะประเมิน 3 ด้าน ได้แก่

5.1 การประเมินประสิทธิภาพด้านความถูกต้องของผลลัพธ์ ความเร็วในการประมวลผล

5.2 การประเมินประสิทธิภาพด้านความครอบคลุมและเพียงพอของคลังคำคำรายอีสานโบราณ มีความเหมาะสมในการนำมาใช้ปรัวรรตคำรายอีสานโบราณ

5.3 การประเมินด้านประโยชน์ที่ได้รับสำหรับการปรัวรรตคำรายอีสานโบราณ

6. แบบประเมินนี้ไม่มีการประเมินประสิทธิภาพด้านความเหมาะสมของการออกแบบหน้าจอหรือส่วนติดต่อกับผู้ใช้ต่าง ๆ (User Interface) และการใช้งานได้ของระบบต้นแบบ (Usability)

ผู้วิจัยขอขอบพระคุณในความอนุเคราะห์ของผู้ประเมินทุกท่าน ที่เสียสละเวลาในการประเมินผลการทำงานในการแปลงคำปรัวรรตของอักษรไทยน้อยเป็นคำอ่านภาษาอีสาน มา ณ โอกาสนี้

ด้วยความเคารพเป็นอย่างสูง
นางสาวจินตนา พลศรี
นักวิจัย

กรุณาทำเครื่องหมาย ✓ ลงในช่องระดับประสิทธิภาพที่ตรงกับความคิดเห็นของท่าน

ข้อ	รายการประเมิน	ระดับประสิทธิภาพที่ได้รับ				
		มากที่สุด	มาก	พอใช้	น้อย	ควรปรับปรุง
1. ความถูกต้องของผลลัพธ์ ความเร็วในการประมวลผล						
1.1	ระดับความถูกต้องของผลลัพธ์มีประสิทธิภาพที่ยอมรับได้					
1.2	ความเร็วในการประมวลผลมีประสิทธิภาพที่ยอมรับได้					
2. ความครอบคลุมและเพียงพอของคลังคำตำรายาอีสานโบราณ มีความเหมาะสมในการนำมาใช้ปริวรรตตำรายาอีสานโบราณ						
2.1	คำศัพท์ครอบคลุมคำศัพท์ด้านตำรายาอีสานโบราณ (ชื่อยา/โรค อาการของโรค เกษษัตถุ วิธีเตรียมยา วิธีใช้ยา หรืออื่นๆ ที่เกี่ยวข้อง)					
2.2	จำนวนคำศัพท์มีเพียงพอสำหรับปริวรรตตำรายาอีสานโบราณได้					
2.3	การสะกดคำอ่านภาษาอีสานมีความเหมาะสม					
3. ประโยชน์ที่ได้รับสำหรับการปริวรรตตำรายาอีสานโบราณ						
3.1	ช่วยให้การปริวรรตตำรายาอีสานโบราณทำได้เร็วขึ้น					
3.2	ช่วยให้การปริวรรตตำรายาอีสานโบราณมีความถูกต้องมากขึ้น					
3.3	ช่วยให้การทำคำอ่านภาษาอีสานในตำรายาเป็นมาตรฐานเดียวกัน					
3.4	ช่วยให้ผู้ที่อ่านอักษรไทยน้อยได้ แต่ไม่มีพื้นฐานความรู้ด้านตำรายาอีสานโบราณสามารถปริวรรตตำรายาอีสานโบราณได้					
3.5	ช่วยลดขั้นตอนและเวลาในการปริวรรตตำรายาอีสานโบราณที่บันทึกในโบราณด้วยอักษรไทยน้อย					

ข้อเสนอแนะเพิ่มเติม

.....

.....

.....

.....



ประวัติผู้เขียน

นางสาวจินตนา พลศรี เกิดเมื่อวันที่ 27 กรกฎาคม พ.ศ. 2514 สำเร็จการศึกษาระดับปริญญาตรีในปีการศึกษา 2537 สาขาวิชาคอมพิวเตอร์ศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏรำไพพรรณี และสำเร็จระดับปริญญาโท สาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ปีการศึกษา 2544 ปัจจุบันรับราชการในตำแหน่งอาจารย์ ประจำสาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ



มหาวิทยาลัยเทคโนโลยีสุรนารี