

การพัฒนาระบบการถ่ายถอดอักษรสำหรับตัวบทเบรลล์ไทยปนอังกฤษ  
เป็นอักษรไทยและอังกฤษปกติโดยใช้วิธีการแบบผสม

นายวิรัช อำพรไพบูลย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรดุษฎีบัณฑิต

สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์

คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2559

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A DEVELOPMENT OF TRANSLITERATION SYSTEM FOR MIXED THAI AND ENGLISH  
BRAILLE TEXTS INTO THAI AND ENGLISH PRINTS USING A HYBRID APPROACH

Mr. Weerachai Umpornpaiboon

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy Program in Linguistics

Department of Linguistics

Faculty of Arts

Chulalongkorn University

Academic Year 2016

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การพัฒนากระบวนการถ่ายถอดอักษรสำหรับตัวบทเบรลล์ไทย  
ปนอังกฤษเป็นอักษรไทยและอังกฤษปกติโดยใช้วิธีการ  
แบบผสม

โดย

นายวีรชัย อำพรไพบูลย์

สาขาวิชา

ภาษาศาสตร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล

---

คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาคุษฎีบัณฑิต

.....คณบดีคณะอักษรศาสตร์

(รองศาสตราจารย์ ดร.กึ่งกาญจนา เทพกาญจนา)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ดร.เทพชัย ทรัพย์นิจิ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.พิทยาวัฒน์ พิทยาภรณ์)

.....กรรมการ

(อาจารย์ ดร.วรรณชัย คำภิระ)

.....กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร.วีระแมน นิยมพล)

วิรัชย์ อำพรไพบูลย์ : การพัฒนาระบบการถ่ายถอดอักษรสำหรับตัวบทเบรลล์ไทยปน  
อังกฤษเป็นอักษรไทยและอังกฤษปกติโดยใช้วิธีการแบบผสม. (A DEVELOPMENT OF  
TRANSLITERATION SYSTEM FOR MIXED THAI AND ENGLISH BRAILLE TEXTS  
INTO THAI AND ENGLISH PRINTS USING A HYBRID APPROACH) อ.ที่ปรึกษา  
วิทยานิพนธ์หลัก : รศ.ดร.วิโรจน์ อรุณมานะกุล, 352 หน้า

วิทยานิพนธ์นี้ศึกษาระบบอักษรเบรลล์ไทยโดยวิเคราะห์อักขระในเบรลล์ไทยและเปรียบเทียบ  
อักขระเบรลล์ไทยกับเบรลล์อังกฤษและเบรลล์ญี่ปุ่น เนื่องจากผู้วิจัยสันนิษฐานว่าอักขระเบรลล์ไทย  
พื้นฐานนำมาจากอักขระในสองภาษานี้ ผลการวิจัยแสดงให้เห็นว่า ในอักขระเบรลล์ไทยมีสระพื้นฐาน  
5 รูป ที่มีรูปและเสียงตรงกับอักขระเบรลล์ญี่ปุ่น และมีพยัญชนะพื้นฐาน 15 รูป ที่มีรูปและเสียง  
ใกล้เคียงกับอักขระเบรลล์อังกฤษ โดยมีการสร้างอักขระขึ้นเพิ่มเติมจากอักขระพื้นฐานเหล่านี้ด้วยการ  
ปรับเปลี่ยนจุดภายในเซลล์ และการเพิ่มอักขระเบรลล์รูปแบบเฉพาะหน้าหรือหลังอักขระที่ใช้เป็นฐาน

นอกจากนี้ วิทยานิพนธ์นี้ยังได้พัฒนาโปรแกรมสำหรับถ่ายถอดอักษรเบรลล์ไทยปนอังกฤษ  
เป็นอักษรปกติ และเปรียบเทียบประสิทธิภาพของระบบที่ใช้วิธีการที่แตกต่างกัน 3 วิธี ได้แก่ การใช้กฎ  
การใช้แบบจำลองเอ็นแกรม และวิธีการแบบผสม โดยทดสอบกับข้อความเบรลล์ 5 ประเภท คือ  
ข้อความเบรลล์ไทย ข้อความเบรลล์อังกฤษรูปเต็ม ข้อความเบรลล์อังกฤษรูปย่อ ข้อความเบรลล์ไทยปน  
อังกฤษรูปเต็ม และข้อความเบรลล์ไทยปนอังกฤษรูปย่อ จากการทดลองพบว่า วิธีการแบบผสมถ่ายถอด  
อักษรได้ถูกต้องมากที่สุดสำหรับข้อความเบรลล์ 4 ประเภท ได้แก่ ข้อความเบรลล์ไทย 99.73 %  
ข้อความเบรลล์อังกฤษรูปเต็ม 100 % ข้อความเบรลล์อังกฤษรูปย่อ 99.92 % และข้อความเบรลล์ไทย  
ปนอังกฤษรูปย่อ 98.69% นอกจากนี้ผลการทดลองยังแสดงให้เห็นว่า วิธีการถ่ายถอดอักษรทั้ง 3 วิธี  
ให้ผลความถูกต้องไม่แตกต่างกันมากนัก

ภาควิชา ..... ภาษาศาสตร์ ..... ลายมือชื่อนิติศ .....  
สาขาวิชา ..... ภาษาศาสตร์ ..... ลายมือชื่อ อ.ที่ปรึกษาหลัก .....  
ปีการศึกษา ..... 2559 .....

##5680506322 : MAJOR LINGUISTICS

KEY WORD : BRAILLE TRANSLITERATION / THAI BRAILLE / ENGLISH BRAILLE /  
RULED-BASED MODEL / N-GRAM MODEL / HYBRID MODEL

WEERACHAI UMPORNPAIBOON : A DEVELOPMENT OF TRANSLITERATION  
SYSTEM FOR MIXED THAI AND ENGLISH BRAILLE TEXTS INTO THAI AND  
ENGLISH PRINTS USING A HYBRID APPROACH. ADVISOR : ASSOC. PROF.  
WIROTE AROONMANAKUN, Ph.D., 352 pp.

This dissertation aims to investigate the Thai Braille writing system, which is believed to be influenced by English and Japanese braille. The findings show that Japanese Braille and Thai Braille share five vowels similar in form and sound, while English Braille and Thai Braille share fifteen consonants with similar sounds. These shared Braille forms are basic characters, which are then extended into other Braille characters representing vowels, consonants, and tones in Thai. The extensions are made through the alternation of dots within the cell or the addition of a preceding or succeeding Braille cell.

Another objective of this dissertation is to develop a program to transliterate Thai mixed with English Braille into print and compare the performances of the systems based on three methods: the rule-based, the N-gram and the hybrid. The systems were tested with five text categories, namely the Thai Braille, the uncontracted Braille, the contracted Braille, the Thai Braille mixed with uncontracted Braille and the Thai mixed with contracted Braille texts. The results reveal that the hybrid method achieves the highest accuracies for the four text categories. The results are as follows: 99.73 % for the Thai Braille texts, 100% for the uncontracted English Braille texts, 99.92% for the contracted English Braille texts and 98.69% for the Thai mixed with contracted English Braille texts. The study also shows that the three methods do not make significant differences in level of performance.

Department : Linguistics ..... Student's Signature .....

Field of Study : Linguistics ..... Advisor's Signature .....

Academic Year : 2016 .....

## กิติกรรมประกาศ

ในการทำวิจัยครั้งนี้ ข้าพเจ้าได้แรงบันดาลใจจาก มิสเจนวีฟ คอลฟิลด์ ผู้ก่อกำเนิดการศึกษาสำหรับคนตาบอดไทย ข้าพเจ้าจึงขอกราบขอบพระคุณท่านอย่างสุดซึ้งที่ทำให้งานวิจัยชิ้นนี้เกิดขึ้นได้

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้ความเมตตาประสิทธิ์ประสาทความรู้ทางด้านภาษาศาสตร์และการเขียนโปรแกรมคอมพิวเตอร์ซึ่งเป็นเครื่องมือที่สำคัญต่อการวิจัยครั้งนี้ อีกทั้งให้คำแนะนำอันเป็นประโยชน์ต่อการวิจัยอย่างสม่ำเสมอตลอดมา ทำให้งานวิจัยนี้สำเร็จลุล่วงไปได้

ข้าพเจ้าขอกราบขอบพระคุณ ดร.เทพชัย ทรัพย์นิธิ ผู้ช่วยศาสตราจารย์ ดร.พิทยาวัฒน์ พิทยาภรณ์ ดร.วรรณชัย คำภีระ และ ดร.วีระแมน นิยมพล คณะกรรมการสอบวิทยานิพนธ์ ที่ได้เมตตาให้คำแนะนำอันเป็นประโยชน์ ทำให้ข้าพเจ้าเกิดมุมมองที่ลึกซึ้งมากยิ่งขึ้นเกี่ยวกับงานวิจัยนี้

ข้าพเจ้าขอกราบขอบพระคุณ ศาสตราจารย์ ดร.อมรา ประสิทธิ์รัฐสินธุ์ ศาสตราจารย์ ดร.ธีระพันธ์ เหลืองทองคำ และ ผู้ช่วยศาสตราจารย์ยุพา ชวนปรีชา ที่ได้เมตตาให้ความรู้และคำแนะนำอันเป็นประโยชน์ต่อการวิจัยครั้งนี้ และขอกราบขอบพระคุณคณาจารย์ภาควิชาภาษาศาสตร์ทุกท่าน ที่ได้ประสิทธิ์ประสาทความรู้ และทักษะในการวิเคราะห์ทางภาษาศาสตร์อันเป็นรากฐานของการวิจัยแก่ข้าพเจ้า

ขอขอบคุณ คุณสุประภา อ่ำพรไพบูลย์ ผู้ช่วยเก็บข้อมูลและจัดพิมพ์วิทยานิพนธ์ฉบับนี้ คุณศุภเดช อันจรสวิชัย ที่ได้ให้คำแนะนำเกี่ยวกับการเขียนโปรแกรมการประมวลผลภาษาธรรมชาติ ศูนย์เทคโนโลยีการศึกษาเพื่อคนตาบอดที่ได้อนุญาตให้ใช้ข้อมูลอักษรเบรลล์ในการวิจัย และเจ้าหน้าที่เลขานุการภาควิชาภาษาศาสตร์ทุกท่านที่ให้ความช่วยเหลือเกื้อกูลแก่ข้าพเจ้าด้วยดีตลอดมา

ท้ายที่สุด ข้าพเจ้าขอกราบขอบพระคุณจุฬาลงกรณ์มหาวิทยาลัย ที่เป็นแหล่งประสิทธิ์ประสาทวิชา อันมีค่ายิ่งต่อข้าพเจ้า ทำให้ข้าพเจ้ามีชีวิตที่มีคุณค่าราบเท่าถึงทุกวันนี้

# สารบัญ

หน้า

กิตติกรรมประกาศ .....	ฉ
สารบัญ .....	ช
สารบัญตาราง .....	ฅ
สารบัญภาพ .....	ด
บทที่ 1 บทนำ .....	1
1.1 ความเป็นมาและความสำคัญของปัญหา .....	1
1.2 วัตถุประสงค์.....	4
1.3 สมมุติฐาน.....	4
1.4 เครื่องมือที่ใช้ในการวิจัย.....	5
1.5 ขอบเขตการวิจัย.....	5
1.6 ข้อมูลที่ใช้ในการวิจัย.....	5
1.7 ขั้นตอนการวิจัย .....	6
1.8 ประโยชน์ที่คาดว่าจะได้รับ .....	6
1.9 โครงสร้างวิทยานิพนธ์ .....	6
1.10 รายการนิยามศัพท์ที่เกี่ยวข้อง .....	7
บทที่ 2 ทบทวนวรรณกรรม.....	9
2.1 ความรู้ทั่วไปเกี่ยวกับอักษรเบรลล์ .....	9
2.1.1 อักษรเบรลล์อังกฤษ.....	10
2.1.2 อักษรเบรลล์ไทย .....	26
2.1.3 ตัวเลขเบรลล์ .....	31
2.2 การจำแนกภาษาด้วยคอมพิวเตอร์ .....	34
2.3 การถ่ายถอดอักษร .....	40
2.3.1 แนวคิดทั่วไปเกี่ยวกับการถ่ายถอดอักษร .....	40
2.3.2 แนวทางการถ่ายถอดอักษรด้วยคอมพิวเตอร์ .....	41
2.4 งานวิจัยที่เกี่ยวข้องกับการถ่ายถอดอักษรเบรลล์ไทยและเบรลล์อังกฤษ.....	53
2.4.1 การออกแบบและพัฒนาระบบถ่ายถอดอักษรไปกลับอังกฤษผสมไทยเป็นเบรลล์ ...	53

2.4.2 การใช้เอ็นแกรมช่วยในการเพิ่มประสิทธิภาพการตัดสินใจถ่ายถอดอักษรเบรลล์ ในกรณีที่มีคำควบกล้ำ สระประสมและอักษรเบรลล์ 2 เซลล์ .....	58
2.5 โปรแกรม Duxbury Braille Translator.....	61
บทที่ 3 ระบบหน่วยอักขระเบรลล์อังกฤษและเบรลล์ไทย.....	63
3.1 ระบบหน่วยอักขระเบรลล์อังกฤษ .....	63
3.1.1 รูปแบบของหน่วยอักขระเบรลล์อังกฤษ.....	63
3.1.2 รูปแบบของตัวเลขเบรลล์.....	65
3.1.3 รูปแบบของเครื่องหมายในเบรลล์อังกฤษ .....	66
3.1.4 อักษรวิธีในเบรลล์อังกฤษ.....	67
3.2 ระบบหน่วยอักขระเบรลล์ไทย.....	67
3.2.1 กรอบการวิเคราะห์ระบบหน่วยอักขระเบรลล์ไทย.....	67
3.2.2 รูปแบบของหน่วยอักขระเบรลล์ไทย .....	68
บทที่ 4 การออกแบบและพัฒนาระบบการถ่ายถอดเบรลล์ไทยปนอังกฤษโดยใช้กฎ .....	85
4.1 แนวคิดที่ใช้ในการออกแบบและพัฒนาระบบการถ่ายถอดเบรลล์ไทย.....	85
4.1.1 ความไม่สอดคล้องระหว่างเบรลล์ไทยและไทยปกติ.....	86
4.1.2 ความกำกวมของเบรลล์เซลล์เดี่ยว .....	89
4.1.3 ความกำกวมของเบรลล์เซลล์คู่.....	93
4.2 แนวคิดที่ใช้ในการออกแบบและพัฒนาระบบการถ่ายถอดเบรลล์อังกฤษ.....	95
4.2.1 เบรลล์อังกฤษรูปเต็ม เบรลล์อังกฤษรูปย่อ และคอมพิวเตอร์เบรลล์ .....	96
4.2.2 แนวคิดที่ใช้ในการถ่ายถอดเบรลล์อังกฤษรูปย่อ .....	99
4.3 แนวคิดที่ใช้ในการระบุภาษาในอักษรเบรลล์ .....	109
4.3.1 การจำแนกภาษาโดยใช้สายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา.....	109
4.3.2 การจำแนกภาษาโดยอาศัยบริบทแวดล้อม.....	110
4.3.3 การจำแนกภาษาโดยใช้ความยาวของสายอักขระ .....	110
บทที่ 5 ระบบการถ่ายถอดเบรลล์ไทยปนอังกฤษโดยใช้กฎ .....	111
5.1 ข้อมูลที่ใช้ในงานวิจัย .....	111
5.1.1 ข้อมูลที่ใช้วิเคราะห์เพื่อสร้างระบบการถ่ายถอดอักษร.....	112
5.1.2 ข้อมูลที่ใช้ในการทดสอบระบบ .....	134
5.2 ระบบการถ่ายถอดอักษรโดยใช้กฎ.....	136
5.2.1 การระบุภาษา.....	138



5.2.2 การถ่ายถอดเบรลล์ไทย .....	139
5.2.3 การถ่ายถอดเบรลล์อังกฤษ.....	154
5.3 การทดสอบระบบการถ่ายถอดอักษร โดยใช้กฎ.....	172
5.3.1 การทดสอบระบบกับข้อความเบรลล์ไม่ปนภาษา.....	175
5.3.2 การทดสอบระบบกับข้อความเบรลล์ปนภาษา .....	176
5.4 อภิปรายผลการทดสอบระบบ .....	178
5.4.1 การระบุภาษา.....	178
5.4.2 การถ่ายถอดอักษร .....	179
บทที่ 6 ระบบการถ่ายถอดเบรลล์ไทยปนอังกฤษโดยใช้แบบจำลองเอ็นแกรม.....	185
6.1 การจัดทำข้อมูลเพื่อใช้กับระบบถ่ายถอดอักษร .....	185
6.1.1 การจัดทำข้อมูลสำหรับการถ่ายถอดเบรลล์ไทย .....	185
6.1.2 การจัดทำข้อมูลสำหรับการถ่ายถอดเบรลล์อังกฤษ .....	189
6.2 ระบบการถ่ายถอดอักษร โดยใช้แบบจำลองเอ็นแกรม .....	198
6.2.1 การสร้างชุดสายอักขระสำหรับภาษาไทยและภาษาอังกฤษ .....	202
6.2.2 การหาผลลัพธ์สำหรับเบรลล์ไทยปนอังกฤษรูปย่อ .....	206
6.2.3 การหาค่าสถิติสูงสุดของสายอักขระ โดยใช้แบบจำลอง ไลตรแกรมตัวอักษร.....	207
6.3 การทดสอบระบบการถ่ายถอดอักษร โดยใช้แบบจำลองเอ็นแกรม .....	208
6.3.1 ผลการทดสอบระบบกับข้อความเบรลล์ไม่ปนภาษา.....	208
6.3.2 การทดสอบระบบกับข้อความเบรลล์ปนภาษา .....	209
6.4 อภิปรายผลการทดสอบระบบ .....	211
6.4.1 การระบุภาษา.....	211
6.4.2 การถ่ายถอดอักษร .....	213
บทที่ 7 การออกแบบและพัฒนาการถ่ายถอดโดยใช้วิธีการแบบผสม.....	218
7.1 ข้อมูลที่ใช้ในงานวิจัย .....	218
7.1.1 ข้อมูลที่ใช้กับระบบถ่ายถอดอักษร .....	218
7.1.2 ข้อมูลที่ใช้ในการทดสอบระบบ .....	220
7.2 ระบบการถ่ายถอดอักษรโดยใช้วิธีการแบบผสม .....	220
7.2.1 การระบุภาษา.....	222
7.2.2 การสร้างชุดสายอักขระสำหรับภาษาไทย.....	223
7.2.3 การสร้างชุดสายอักขระสำหรับภาษาอังกฤษ.....	226

7.2.4 การหาค่าสถิติสูงสุดด้วยแบบจำลองเอ็นแกรม .....	227
7.3 การทดสอบระบบการถ่ายถอดอักษรโดยใช้วิธีการแบบผสม .....	227
7.3.1 ผลการทดสอบระบบกับข้อความเบรลล์ไม่ปนภาษา .....	227
7.3.2 การทดสอบระบบกับข้อความเบรลล์ปนภาษา .....	229
7.4 อภิปรายผลการทดสอบระบบ .....	230
7.4.1 การระบุภาษา .....	230
7.4.2 การถ่ายถอดอักษร .....	231
7.5 ผลการเปรียบเทียบวิธีการถ่ายถอดอักษรทั้ง 3 วิธี .....	232
7.5.1 ผลการเปรียบเทียบการระบุภาษา .....	232
7.5.2 ผลการเปรียบเทียบการถ่ายถอดอักษร .....	233
บทที่ 8 สรุปผล อภิปรายผล และข้อเสนอแนะ .....	236
8.1 สรุปผล .....	236
8.1.1 หน่วยอักขระเบรลล์ไทยและหน่วยอักขระเบรลล์อังกฤษ .....	237
8.1.2 ระบบการถ่ายถอดเบรลล์ไทยปนอังกฤษ .....	239
8.2 อภิปรายผล .....	241
8.3 ข้อเสนอแนะ .....	243
รายการอ้างอิง .....	248
ภาคผนวก .....	253
ประวัติผู้เขียนวิทยานิพนธ์ .....	321

## สารบัญตาราง

หน้า

ตารางที่ 2.1 ตัวอักษรเบรลล์อังกฤษ .....	11
ตารางที่ 2.2 การใช้ตัวย่ออักษร โศด .....	13
ตารางที่ 2.3 รายการตัวย่ออักษร โศด .....	14
ตารางที่ 2.4 วิธีการใช้ตัวย่อรูปเต็มคงรูป .....	15
ตารางที่ 2.5 รายการตัวย่อรูปเต็มคงรูป.....	16
ตารางที่ 2.6 รายการตัวย่อรูปเต็มกลายรูป .....	16
ตารางที่ 2.7 รายการตัวย่อประกอบคำ.....	17
ตารางที่ 2.8 วิธีการใช้ตัวย่อประกอบคำ.....	17
ตารางที่ 2.9 รายการตัวย่อกลางคำหรือท้ายคำ.....	18
ตารางที่ 2.10 วิธีการใช้ตัวย่อกลางคำหรือท้ายคำ.....	18
ตารางที่ 2.11 รายการตัวย่อเบรลล์ต่ำย่อคำ โศด .....	19
ตารางที่ 2.12 รายการตัวย่อเบรลล์ต่ำที่ย่อคำ โศดพิเศษ .....	19
ตารางที่ 2.13 รายการตัวย่อเบรลล์ต่ำย่อส่วนของคำ.....	20
ตารางที่ 2.14 รายการตัวย่อหน่วยหลัก.....	21
ตารางที่ 2.15 การใช้ตัวย่อหน่วยหลัก .....	21
ตารางที่ 2.16 รายการตัวย่อหน่วยเสริม .....	22
ตารางที่ 2.17 รายการคำย่อเบรลล์ .....	24
ตารางที่ 2.18 การใช้คำย่ออักษรเบรลล์ .....	25
ตารางที่ 2.19 พยัญชนะในเบรลล์ไทย.....	27
ตารางที่ 2.20 ตัวอักษรเบรลล์ไทยพื้นฐาน 15 ตัว .....	28
ตารางที่ 2.21 สระและวรรณยุกต์ในเบรลล์ไทย .....	29
ตารางที่ 2.22 ตัวเลขในอักษรเบรลล์ .....	32
ตารางที่ 2.23 เครื่องหมายที่ปรากฏกับตัวเลข.....	33
ตารางที่ 2.24 เบรลล์เซลล์เดี่ยวที่กำวม.....	56
ตารางที่ 2.25 เบรลล์เซลล์คู่ที่กำวม .....	58

ตารางที่ 2.26 ตัวอย่างข้อความไทยที่โปรแกรม DUXBURY BRAILLE TRANSLATOR ถ่ายถอดผิด	61
ตารางที่ 2.27 ตัวอย่างข้อความไทยปนอังกฤษที่โปรแกรม DUXBURY BRAILLE TRANSLATOR ถ่ายถอดผิด	62
ตารางที่ 3.1 หน่วยอักขระเบรลล์อังกฤษพื้นฐาน	64
ตารางที่ 3.2 หน่วยอักขระเบรลล์อังกฤษดัดแปลงที่ได้จากการเติม ‘’ ภายในเซลล์	64
ตารางที่ 3.3 หน่วยอักขระเบรลล์อังกฤษดัดแปลงที่ได้จากการเติม ‘-’ ภายในเซลล์	65
ตารางที่ 3.4 ตัวเลขเบรลล์ 1 - 0	66
ตารางที่ 3.5 เครื่องหมายในเบรลล์อังกฤษ	66
ตารางที่ 3.6 อักขระเบรลล์ที่มีรูปซ้ำกันระหว่างภาษาไทยและญี่ปุ่น หรือไทยและอังกฤษ	69
ตารางที่ 3.7 อักขระในเบรลล์ญี่ปุ่นและเบรลล์ไทยที่อยู่ในกลุ่มเบรลล์อังกฤษพื้นฐาน	71
ตารางที่ 3.8 รูปของสระในเบรลล์ไทยและเบรลล์ญี่ปุ่นที่มีเสียงใกล้เคียงกัน	72
ตารางที่ 3.9 พยัญชนะเบรลล์ไทยที่มีรูปและเสียงตรงกับเบรลล์อังกฤษ	72
ตารางที่ 3.10 การสร้างสระขึ้นเพิ่มเติม โดยการเติม ‘;’ (จุด 6)	74
ตารางที่ 3.11 การสร้างอักษร โดยการเลื่อนตำแหน่งจุดเบรลล์ให้ต่ำลง	75
ตารางที่ 3.12 การสร้างสระจาก :: (=๖)	75
ตารางที่ 3.13 อักขระที่เกิดจากการปรับเปลี่ยนจุดเบรลล์แบบอื่น ๆ	76
ตารางที่ 3.14 การสร้างสระสั้น โดยการเพิ่มอักขระ ‘:’ (จุด 1)	77
ตารางที่ 3.15 การสร้างสระเพิ่มเติม โดยการเพิ่มอักขระ ‘::’ (จุด 2)	78
ตารางที่ 3.16 พยัญชนะเบรลล์ไทยที่มีรูปและเสียงตรงกับเบรลล์อังกฤษจำแนกตามหมู่อักษร	79
ตารางที่ 3.17 พยัญชนะเบรลล์ไทยที่เกิดจากการดัดแปลงพยัญชนะเบรลล์อังกฤษ	79
ตารางที่ 3.18 การสร้างพยัญชนะใหม่ที่มีหมู่อักษรต่างจากพยัญชนะฐาน	80
ตารางที่ 3.19 การสร้างพยัญชนะใหม่ที่มีหมู่ตัวอักษรและเสียงตรงกับพยัญชนะฐาน	82
ตารางที่ 5.1 จำนวนสายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา	114
ตารางที่ 5.2 เครื่องหมายเบรลล์ค้นสายอักขระ	114
ตารางที่ 5.3 คู่เทียบเบรลล์เซลล์เดี่ยวทั่วไป	115
ตารางที่ 5.4 คู่เทียบสระทั่วไป	116
ตารางที่ 5.5 คู่เทียบสระพิเศษ	116
ตารางที่ 5.6 คู่เทียบสระหน้า	117

ตารางที่ 5.7	คู่มือเทียบวรรณยุกต์.....	117
ตารางที่ 5.8	คู่มือเทียบเบรลล์เซลล์คู่ทั่วไป.....	118
ตารางที่ 5.9	คู่มือเทียบสระประสม.....	119
ตารางที่ 5.10	คู่มือเทียบเบรลล์เซลล์เดี่ยวที่กำกวมในตำแหน่งทั่วไป.....	120
ตารางที่ 5.11	คู่มือเทียบเบรลล์ที่กำกวมท้ายสายอักขระ.....	120
ตารางที่ 5.12	คู่มือเทียบสำหรับเบรลล์เซลล์คู่ที่กำกวม.....	121
ตารางที่ 5.13	เซตพยัญชนะตัวสะกดท้ายพยางค์สำหรับการถ่ายถอด ‘:’ ให้เป็น ‘:’.....	122
ตารางที่ 5.14	เซตของพยัญชนะควบกำกวมและพยัญชนะควบไม่กำกวม.....	123
ตารางที่ 5.15	คู่มือเทียบเบรลล์ต้นสายอักขระ.....	126
ตารางที่ 5.16	คู่มือเทียบเครื่องหมายหลังคำ.....	127
ตารางที่ 5.17	คู่มือเทียบตัวย่อต้นคำ.....	128
ตารางที่ 5.18	คู่มือเทียบตัวย่อกลางคำ.....	129
ตารางที่ 5.19	คู่มือเทียบอักษรเบรลล์หลังตัวย่อระดับคำ.....	130
ตารางที่ 5.20	คู่มือเทียบเครื่องหมายที่ใช้คั่นระหว่างคำ.....	131
ตารางที่ 5.21	คู่มือเทียบตัวย่อรูปเต็มคงรูป.....	131
ตารางที่ 5.22	คู่มือเทียบตัวย่อเบรลล์เซลล์เดี่ยวทั่วไป.....	132
ตารางที่ 5.23	คู่มือเทียบตัวเลขเบรลล์.....	133
ตารางที่ 5.24	ปริมาณของข้อความเบรลล์แต่ละประเภทที่นำมาทดสอบระบบ.....	136
ตารางที่ 5.25	การทดสอบประสิทธิภาพของกระบวนการต่าง ๆ ในข้อมูลที่ไม่มีการปนภาษา.....	173
ตารางที่ 5.26	การทดสอบประสิทธิภาพของกระบวนการต่าง ๆ ในข้อมูลปนภาษา.....	174
ตารางที่ 5.27	ผลการระบุภาษาในข้อความเบรลล์ไม่ปนภาษา.....	175
ตารางที่ 5.28	ผลการถ่ายถอดอักขรในข้อความเบรลล์ไม่ปนภาษา.....	176
ตารางที่ 5.29	ผลการระบุภาษาที่ทดสอบกับข้อความเบรลล์ปนภาษา.....	177
ตารางที่ 5.30	ผลการถ่ายถอดอักขรเฉพาะสายอักขระที่ระบุภาษาได้ถูกต้อง.....	177
ตารางที่ 5.31	ผลการถ่ายถอดอักขรด้วยการใช้กฎโดยภาพรวม.....	178
ตารางที่ 6.1	ผลการทดสอบประสิทธิภาพของการระบุภาษาโดยใช้แบบจำลองเอ็นแกรม.....	209
ตารางที่ 6.2	ผลการทดสอบประสิทธิภาพของการถ่ายถอดอักขรโดยใช้แบบจำลองเอ็นแกรม.....	209
ตารางที่ 6.3	ผลการระบุภาษาของระบบถ่ายถอดอักขรที่ทดสอบกับข้อมูลปนภาษา.....	210

ตารางที่ 6.4 ผลการถ่ายถอดอักษรเฉพาะข้อความที่ระบุภาษาถูกต้อง.....	210
ตารางที่ 6.5 ผลการถ่ายถอดอักษรด้วยการใช้แบบจำลองเอ็นแกรม โดยภาพรวม .....	211
ตารางที่ 7.1 ผลการทดสอบประสิทธิภาพของการระบุภาษาโดยใช้วิธีการแบบผสม .....	228
ตารางที่ 7.2 ผลการทดสอบประสิทธิภาพของการถ่ายถอดอักษรโดยใช้วิธีการแบบผสม .....	228
ตารางที่ 7.3 ผลการระบุภาษาของระบบที่ทดสอบกับข้อมูลป็นภาษา.....	229
ตารางที่ 7.4 ผลการถ่ายถอดอักษรเฉพาะข้อความที่ระบุภาษาได้ถูกต้อง .....	229
ตารางที่ 7.5 ผลการถ่ายถอดอักษรด้วยการใช้วิธีการแบบผสม โดยภาพรวม .....	230
ตารางที่ 7.6 ประสิทธิภาพของการระบุภาษาด้วยวิธีการทั้ง 3 วิธี .....	232
ตารางที่ 7.7 ประสิทธิภาพของการถ่ายถอดอักษรด้วยวิธีการทั้ง 3 วิธี.....	233

## สารบัญภาพ

	หน้า
ภาพที่ 2.1 การปรากฏของจุดในอักษรเบรลล์ .....	10
ภาพที่ 2.2 วิธีการ WEIGHTED FINITE-STATE TRANSDUCER (ที่มา KNIGHT AND GRAEHL,1997).....	48
ภาพที่ 2.3 โครงสร้างแบ็กพรอพาทเกชันนิวโรลเน็ตเวิร์กที่ใช้ในการเรียนรู้คำไทย .....	50
ภาพที่ 4.1 การจัดเรียงสระพิเศษในอักษรเบรลล์และในอักษรปกติ .....	86
ภาพที่ 4.2 การจัดเรียงสระหน้าในอักษรเบรลล์และในอักษรปกติ.....	87
ภาพที่ 5.1 ภาพรวมของระบบถ่ายถอดเบรลล์ไทยปนอังกฤษ โดยใช้กฎ .....	137
ภาพที่ 5.2 กระบวนการถ่ายถอดเบรลล์ไทยโดยใช้กฎ .....	140
ภาพที่ 5.3 กระบวนการถ่ายถอดเบรลล์อังกฤษโดยใช้กฎ .....	155
ภาพที่ 5.4 ขั้นตอนในการถ่ายถอดเบรลล์อังกฤษรูปย่อ .....	166
ภาพที่ 6.1 ภาพรวมของระบบถ่ายถอดเบรลล์ไทยปนอังกฤษรูปเต็ม โดยใช้แบบจำลองเอ็นแกรม 199 .....	
ภาพที่ 6.2 ภาพรวมของระบบถ่ายถอดเบรลล์ไทยปนอังกฤษรูปย่อ โดยใช้แบบจำลองเอ็นแกรม 201 .....	
ภาพที่ 7.1 ภาพรวมของระบบถ่ายถอดเบรลล์ไทยปนอังกฤษ โดยใช้วิธีการแบบผสม .....	221

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ผู้พิการทางสายตามีความจำเป็นที่ต้องรับข้อมูลจากสื่อต่าง ๆ เช่นเดียวกับคนสายตปกติทั่วไป แต่เนื่องจากไม่สามารถจะรับรู้อักษรผ่านการมองเห็นได้ จึงจำเป็นที่จะต้องมียระบบอักษรเบรลล์ ซึ่งเป็นระบบที่แต่ละเซลล์ประกอบด้วยจุดนูนไม่เกินหกจุด (::) มาเรียงต่อกัน โดยที่การปรากฏของจุดในตำแหน่งที่ต่างกันใช้แสดงอักษรที่ต่างกัน ทำให้คนที่มองไม่เห็นสามารถรับรู้ข้อมูลได้โดยการสัมผัสด้วยปลายนิ้วมือ

เป็นที่ทราบกันดีว่า อักษรเบรลล์มีต้นกำเนิดมาจากประเทศฝรั่งเศสซึ่งใช้ระบบตัวอักษรร่วมกับภาษาอื่น ๆ ที่ใช้อักษรโรมัน และมีการศึกษาวิจัยเกี่ยวกับตัวอักษรเบรลล์ในภาษาเหล่านี้อย่างกว้างขวางโดยมีคำอธิบายเกี่ยวกับการสร้างตัวอักษรเบรลล์อย่างเป็นระบบ ตัวอักษรเบรลล์ที่ใช้อักษรโรมันเกิดจากการนำเอาอักษรเบรลล์ 10 ตัวแรกคือ ‘::’ - ‘:::’ (=a-j) มาใช้เป็นอักษรหลักเพื่อสร้างอักษรเบรลล์ตัวอื่น ๆ ในระบบ (Englebretson, R. 2009: 3)

ในการเรียนการสอนเบรลล์ไทยนั้น มีตำราสอนอ่านเขียนเบรลล์ไทยเกิดขึ้นบ้างแต่ตำราเหล่านี้เน้นการท่องจำรหัสอักษรเบรลล์แต่ละตัว ในปัจจุบันยังไม่มีเอกสารหรืองานวิจัยชิ้นใดที่อธิบายเบรลล์ไทยอย่างเป็นระบบ แต่ในความเป็นจริงแล้วในระบบเบรลล์ไทยเองนั้นอักษรตัวหนึ่งก็น่าจะมีความสัมพันธ์กับอักษรตัวอื่นด้วยในลักษณะที่คล้ายคลึงกับระบบในภาษาอื่น เช่น ‘::’ ใช้แทน “ส” ‘:::’ ใช้แทน “ซ” พยัญชนะสองตัวนี้มีเสียงคล้ายคลึงกัน ต่างกันตรงที่ “ส” เป็นอักษรสูง ในขณะที่ “ซ” เป็นอักษรต่ำ และผู้วิจัยยังสังเกตเห็นว่าเบรลล์ไทยก็มีความสัมพันธ์กับเบรลล์อังกฤษด้วย

เหตุที่ว่าเบรลล์แต่ละเซลล์เกิดจากการเรียงต่อของจุดไม่เกินหกจุด ทำให้อักษรเบรลล์ในภาษาหนึ่งเกิดการซ้ำรูปกับอีกภาษาหนึ่งอย่างหลีกเลี่ยงไม่ได้ หากพิจารณารูปซ้ำของเบรลล์อังกฤษและเบรลล์ไทย จะพบว่า มีรูปซ้ำจำนวนหนึ่งที่ใช้แทนเสียงเดียวกัน หรือเสียงที่ใกล้เคียงกันทั้งสองภาษา เช่น “ร” และ “ส” มีคุณสมบัติทางเสียงใกล้เคียงกันมากจึงใช้ ‘::’ แทนตัวอักษรทั้งสองตัว

หลักฐานข้างต้นแสดงให้เห็นว่าเบรลล์ไทยมีรูปที่ปรากฏอย่างเป็นระบบ จึงน่าจะสนใจว่าเบรลล์ไทยมีความสัมพันธ์กับเบรลล์อังกฤษหรือไม่อย่างไรและเบรลล์ไทยแต่ละตัวมีความสัมพันธ์



กันหรือไม่ ในลักษณะใด วิทยานิพนธ์ฉบับนี้จึงมุ่งที่จะหาคำตอบในประเด็นเหล่านี้ ซึ่งเป็นประเด็นพื้นฐานทางภาษาศาสตร์สำหรับการเรียนการสอนอักษรเบรลล์ไทย

นอกจากนี้ พัฒนาการของเทคโนโลยีทำให้มีการใช้อักษรเบรลล์สำหรับภาษาต่างๆ ในคอมพิวเตอร์อย่างแพร่หลาย แต่อักษรเบรลล์มีการเข้ารหัสระหว่างภาษาจึงยากที่เครื่องคอมพิวเตอร์จะบอกได้ว่ากำลังทำงานกับอักษรเบรลล์ของภาษาใดอยู่ แม้ว่าจะมีโปรแกรมอยู่จำนวนหนึ่งที่สามารถถ่ายถอดอักษรเบรลล์เป็นอักษรปกติได้ แต่น่าเสียดายว่าในการถ่ายถอดอักษรนั้น ต้องให้ผู้ใช้กำกับภาษาเพื่อบอกให้คอมพิวเตอร์รู้ว่ากำลังทำงานกับภาษาใด จึงทำให้ในปัจจุบันยังไม่มีโปรแกรมใดที่รองรับการถ่ายถอดข้อความเบรลล์ไทยปนอังกฤษเป็นอักษรปกติโดยอัตโนมัติได้

เพื่อแก้ปัญหาดังกล่าว ได้มีนักวิจัยไทยพยายามเสนอวิธีการจำแนกภาษาของอักษรเบรลล์ ดังที่ปรากฏในงานวิจัยของวิศม์ วงษ์สมาน (2551) ในส่วนหนึ่งของการวิจัย วิศม์ได้ศึกษาการถ่ายถอดอักษรเบรลล์เป็นอักษรปกติ โดยใช้แบบจำลองเอ็นแกรมเข้ามาช่วยในการจำแนกอักษรเบรลล์ของภาษาไทยและภาษาอังกฤษออกจากกันให้ได้ ก่อนที่จะทำการถ่ายถอดแต่ละภาษาเป็นอักษรปกติ วิศม์พบว่าเมื่อมีข้อผิดพลาดที่เกิดจากการระบุภาษาผิด เช่นการถ่ายถอด ‘:: :: :: ::’ ซึ่งควรเป็นคำว่า “ละคร” ในภาษาไทย แต่กลับได้เป็น "laur" ในภาษาอังกฤษ (วิศม์ วงษ์สมาน, 2551: 60)

เป็นที่น่าสังเกตว่าในการวิจัย วิศม์ได้ใช้ข้อมูลอักษรเบรลล์ในการฝึกระบบ ซึ่งก่อให้เกิดความยุ่งยากหลายประการ ประการแรก การใช้ข้อมูลอักษรเบรลล์ฝึกระบบทำให้สิ้นเปลืองเวลาและทรัพยากร เพราะในการใช้คลังข้อมูลอักษรเบรลล์ต้องอาศัยผู้เชี่ยวชาญตรวจสอบข้อมูลอักษรเบรลล์ให้ถูกต้องเสียก่อน มิฉะนั้นแล้วจะทำให้คอมพิวเตอร์ประมวลผลได้ไม่ถูกต้อง

นอกจากนี้การใช้ข้อมูลอักษรเบรลล์มาฝึกระบบ ยังส่งผลให้ไม่สามารถแยกภาษาไทยและอังกฤษออกจากกันได้อย่างเป็นเอกเทศ เนื่องจากอักษรเบรลล์ทั้ง 2 ภาษานี้ต่างก็ใช้ระบบหกจุด ทำให้สายอักขระบางส่วนในคลังข้อมูลภาษาหนึ่งตรงกับสายอักขระในคลังข้อมูลของอีกภาษาหนึ่งได้ หากข้อมูลที่ต้องการถ่ายถอดตรงกับสายอักขระที่ตรงกันดังกล่าว ทำให้ไม่สามารถบ่งชี้ว่าเป็นภาษาใดได้ทันที เมื่อพิจารณาสิ่งที่กล่าวมาข้างต้นจะเห็นได้ว่าคอมพิวเตอร์ไม่สามารถบ่งชี้ว่า ‘:: :: :: ::’ เป็น “ละคร” หรือ “laur” เนื่องจาก ‘:: :: :: ::’ ปรากฏอยู่ในคลังข้อมูลทั้งภาษาไทยและภาษาอังกฤษ

ดังนั้น วิทยานิพนธ์ฉบับนี้จึงมุ่งที่จะนำเสนอวิธีการจำแนกเบรลล์อังกฤษและเบรลล์ไทยออกจากกัน โดยใช้กฎทางภาษาร่วมกับแบบจำลองเอ็นแกรม ที่ผ่านการฝึกระบบด้วยคลังข้อมูลอักษรปกติของทั้ง 2 ภาษา แทนที่จะใช้คลังข้อมูลอักษรเบรลล์โดยตรง

ในการถ่ายถอดเบรลล์ที่มีการปนกันระหว่างภาษาไทยและอังกฤษ นอกจากปัญหาในการจำแนกทั้งสองภาษาออกจากกันแล้ว อักษรเบรลล์ไทยเองก็ยังพบปัญหาความกำกวมในการถ่ายถอดให้เป็นไทยปกติ เหตุที่เป็นเช่นนี้เพราะในเบรลล์ไทยมีการนำเบรลล์หนึ่งเซลล์มาแทนอักขระหลาย

ตัวในการเขียนสระประสม เช่น ‘:’ ใช้แทน “ัว” หรือ ‘:’ ใช้แทน “ื่อ” เป็นต้น ในการเขียนเบรลล์ไทยจะนำสระประสมไปวางไว้หลังพยัญชนะต้น เช่นคำว่า “เรียน” ในเบรลล์ไทยจะเขียนว่า ‘: :: ::’ (=ร-ื่อ-น) คำว่า “เรื่อง” เขียนว่า ‘: :: :: ::’ (=ร-ื่อ-ง) เป็นต้น ปัญหาความกำกวมจะเกิดขึ้นในกรณีที่พยัญชนะควบกล้ำเกิดหน้าสระประสมเหล่านี้ ดังจะเห็นได้จากงานของ วัชรวิมล วังษ์สมาน (2551) ที่ศึกษาการถ่ายทอดเบรลล์ไทยแบบไฟล์ต่อไฟล์โดยใช้กฎทางภาษา และวรวพล ทินกรสูติบุตร (2551) ที่ศึกษาการถ่ายทอดเบรลล์ไทยแบบทันทีทันใดโดยใช้กฎ ทั้งวัชรวิมลและวรวพล พบปัญหาในการถ่ายทอดสระที่ตัวอักษรนำหน้าสองตัวมีรูปคล้ายพยัญชนะควบกล้ำ แต่ไม่ใช่พยัญชนะควบกล้ำเช่น ‘: :: :: :: :: :: :: :: :: :: ::’ (บ-อ-หิ-น-น-อ-ก-ล-แ-ะ) คือ “บ่อหินนอก และ” ถูกถ่ายทอดเป็น “บ่อหินนอกและ” เนื่องจากคอมพิวเตอร์ตีความให้ “กล” เป็นพยัญชนะควบกล้ำ แต่ที่จริง “ก” เป็นตัวสะกดของคำแรก และ “ล” เป็นพยัญชนะต้นของคำถัดมา

ทศวัฒน์ ชุณหวิหะธีระ (2555) ได้พยายามแก้ปัญหาดังกล่าว โดยเสนอให้มีการตัดคำในอักษรเบรลล์ด้วยเทคนิคเอ็นแกรมที่ผ่านการฝึกระบบด้วยคลังข้อมูลอักษรเบรลล์ แล้วนำคำเบรลล์แต่ละคำมาเทียบกับพจนานุกรมเบรลล์เพื่อถ่ายทอดเป็นอักษรปกติ ทศวัฒน์พบว่าวิธีการนี้สามารถแก้ปัญหาพยัญชนะที่กำกวมได้ในระดับหนึ่ง แต่วิธีการนี้ไม่เหมาะสมในการนำมาใช้กับเพิ่มข้อมูลที่มีขนาดใหญ่ เพราะต้องใช้เวลาในการประมวลผลมาก อีกทั้งวิธีการนี้ต้องอาศัยทั้งพจนานุกรมและคลังข้อมูลเบรลล์ที่ต้องสมบูรณ์จึงจะประมวลผลได้ถูกต้อง ซึ่งเป็นไปได้ยากเนื่องจากต้องใช้เวลาและมีค่าใช้จ่ายสูง

ในงานวิจัยนี้ ผู้วิจัยจึงขอเสนอวิธีการถ่ายทอดอักษรเบรลล์ไทยปนอังกฤษโดยใช้คลังข้อความภาษาไทยและภาษาอังกฤษเป็นหลัก ร่วมกับการใช้กฎทางภาษาและความน่าจะเป็นทางสถิติ ซึ่งจะนำเสนอต่อไป

นอกจากนี้ในการใช้เบรลล์ไทยปนอังกฤษนั้น ผู้วิจัยยังพบว่าเบรลล์อังกฤษมีการใช้ใน 2 รูปแบบ คือ เบรลล์อังกฤษรูปเต็ม (English braille grade 1) และ เบรลล์อังกฤษรูปย่อ (English braille grade 2) ซึ่งงานวิจัยการถ่ายทอดเบรลล์ไทยปนอังกฤษที่ผ่านมาศึกษาเฉพาะเบรลล์อังกฤษรูปเต็มเท่านั้น แต่ในความเป็นจริงมีการใช้ภาษาอังกฤษทั้งสองรูปแบบ โดยเฉพาะหากเป็นการจดบันทึกจะนิยมใช้รูปย่อมากกว่า งานวิจัยนี้จึงสนใจศึกษาการถ่ายทอดเบรลล์ไทยปนอังกฤษทั้งรูปเต็มและรูปย่อเพราะจะเป็นประโยชน์ต่อคนตาบอดในการนำไปใช้งานมากยิ่งขึ้น

สำหรับเบรลล์ไทยรูปย่อ (Thai Braille grade 2) ได้มีการคิดค้นโดยกลุ่มคนในสถานศึกษาเพื่อคนตาบอดด้วยหวังว่าจะช่วยให้คนตาบอดไทยจดบันทึกด้วยอักษรเบรลล์ได้สะดวกและรวดเร็ว แต่ยังไม่มีการศึกษาความเหมาะสมในการใช้อักษรเบรลล์ระบบนี้เลย จากการศึกษาเบื้องต้นผู้วิจัยได้ทำการสุ่มตัวอย่างโดยสัมภาษณ์คนตาบอดกลุ่มหนึ่งในสถานศึกษา พบว่าเบรลล์ไทยรูปย่อที่

ประดิษฐ์ขึ้นใหม่มีความซับซ้อนในการใช้เป็นอย่างมาก ตัวอย่างเช่นการนำเบรลล์ 1 เซลล์มาใช้ย่อยส่วนของคำ เช่น ‘::’ แทนส่วนของคำ “-าน” โดยวางสัญลักษณ์ไว้หน้าพยัญชนะต้นของพยางค์ อีกทั้งคำที่ปรากฏในเบรลล์ไทยรูปย่อก็ไม่ได้เกิดจากการศึกษาวิจัยอย่างเป็นระบบ คำย่อหลายคำมีความถี่ในการใช้น้อยมาก เช่น ‘:: :: :: :: ::’ (ศ-ฐ) ใช้แทน “เศรษฐี” ‘:: :: :: ::’ (พพ-) ใช้แทน “พหูพจน์” และ ‘:: :: :: :: ::’ (ป-ชช) ใช้แทน “ประชาชาติ” ฯลฯ เบรลล์ไทยรูปย่อจึงไม่เป็นที่นิยมของกลุ่มคนตาบอดไทยที่ผู้วิจัยทำการสัมภาษณ์ นอกจากนี้ผู้วิจัยได้ทำการสุ่มตัวอย่างเพื่อหาว่ามีการใช้เบรลล์ไทยรูปย่อในสื่อสิ่งพิมพ์อักษรเบรลล์บ้างหรือไม่ โดยสำรวจจากห้องสมุดอักษรเบรลล์หลายแห่ง พบว่ายังไม่มีการใช้อักษรเบรลล์รูปแบบนี้ในเอกสารที่ทำการสุ่มแต่ประการใด ด้วยเหตุที่กล่าวมาข้างต้น ผู้วิจัยจึงไม่นำเบรลล์ไทยรูปย่อมาใส่ไว้ในขอบเขตของงานวิจัยครั้งนี้

## 1.2 วัตถุประสงค์

1. วิเคราะห์ระบบอักษรเบรลล์ไทยและระบบอักษรเบรลล์อังกฤษ
2. พัฒนาระบบการถอดถอดอักษรสำหรับตัวบทเบรลล์ไทยปนอังกฤษ โดยการใช้กฎทางภาษาร่วมกับแบบจำลองเอ็นแกรม (n-gram)

## 1.3 สมมุติฐาน

1. หน่วยอักขระเบรลล์ไทยสามารถจำแนกเป็น 3 ประเภท ได้แก่ หน่วยอักขระเบรลล์ไทยพื้นฐาน หน่วยอักขระเบรลล์ไทยดัดแปลง และหน่วยอักขระเบรลล์ไทยจำเพาะ หน่วยอักขระเบรลล์ไทยพื้นฐานแต่ละตัวมีรูปและเสียงสอดคล้องกับหน่วยอักขระเบรลล์ในภาษาอังกฤษและภาษาญี่ปุ่น หน่วยอักขระเบรลล์ไทยดัดแปลงแต่ละตัวได้มาจากการปรับหน่วยอักขระเบรลล์ไทยพื้นฐานที่มีเสียงตรงกันโดยการกำกับด้วยเครื่องหมาย (glyph) และหน่วยอักขระเบรลล์ไทยจำเพาะเป็นหน่วยอักขระเบรลล์ที่ปรากฏเฉพาะภาษาไทยเท่านั้น
2. หน่วยอักขระเบรลล์อังกฤษสามารถจำแนกเป็น 2 ประเภท ได้แก่ หน่วยอักขระเบรลล์อังกฤษพื้นฐาน และหน่วยอักขระเบรลล์อังกฤษดัดแปลง หน่วยอักขระเบรลล์อังกฤษพื้นฐานประกอบด้วยหน่วยอักขระเบรลล์ 10 ตัวแรกในภาษาอังกฤษ (a - j) และหน่วยอักขระเบรลล์อังกฤษดัดแปลงได้มาจากการกำกับหน่วยอักขระเบรลล์อังกฤษพื้นฐานแต่ละตัวด้วยเครื่องหมาย (glyph)
3. ระบบการถอดถอดอักษรสำหรับตัวบทเบรลล์ไทยปนอังกฤษเป็นอักษรปกติที่อาศัยกฎทางภาษาร่วมกับแบบจำลองเอ็นแกรม จะให้ผลลัพธ์ที่ถูกต้องมากกว่าการใช้วิธีการใดวิธีการหนึ่ง

#### 1.4 เครื่องมือที่ใช้ในการวิจัย

1. เครื่องไมโครคอมพิวเตอร์ส่วนบุคคล
2. โปรแกรมภาษา Perl ของบริษัท Active Perl
3. เครื่อง Braille Display ยี่ห้อ Focus ของบริษัท Freedom Scientific

#### 1.5 ขอบเขตการวิจัย

1. ศึกษาการถ่ายถอดเบรลล์ไทยรูปเต็ม (Thai Braille Grade1) เบรลล์อังกฤษรูปเต็ม(English Braille Grade1) และเบรลล์อังกฤษรูปย่อ (English Braille Grade2) เป็นอักษรปกติ
2. ศึกษาเฉพาะสัญลักษณ์เบรลล์ที่ปรากฏในข้อความทั่วไป ไม่รวมถึงสัญลักษณ์ที่ปรากฏเฉพาะด้าน เช่น คนตรี และคณิตศาสตร์

#### 1.6 ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในงานวิจัยนี้ แบ่งออกเป็น 2 กลุ่ม ได้แก่

##### 1.6.1 ข้อมูลในการวิเคราะห์ระบบตัวอักษรเบรลล์ไทยและระบบตัวอักษรเบรลล์อังกฤษ

1.6.1.1 รายการตัวอักษรเบรลล์ไทย ได้มาจาก ตำราสอนอ่านเขียนอักษรเบรลล์ (สมทรง พันธุ์สุวรรณ, 2538)

1.6.1.2 รายการตัวอักษรเบรลล์อังกฤษ ได้มาจากตำรา Instruction Manual for Braille Transcribing (Risjord, Constance, 2009)

1.6.2.3 รายการตัวอักษรเบรลล์ญี่ปุ่น ได้จากเว็บไซต์การเรียนการสอนอักษรเบรลล์ญี่ปุ่น ([www.jfbkk.or.th/pdf/JL/2014](http://www.jfbkk.or.th/pdf/JL/2014))

##### 1.6.2 ข้อมูลที่ใช้ในระบบการถ่ายถอดอักษรเบรลล์

###### 1.6.2.1 ข้อมูลที่ใช้วิเคราะห์เพื่อสร้างระบบการถ่ายถอดอักษร

ข้อมูลชุดนี้ประกอบด้วย ข้อมูลที่ใช้ในการระบุภาษา ข้อมูลที่ใช้ในการถ่ายถอดเบรลล์ไทย และข้อมูลที่ใช้ในการถ่ายถอดเบรลล์อังกฤษ

###### 1.6.2.2 ข้อมูลที่นำมาใช้ทดสอบระบบ

ข้อมูลชุดนี้ประกอบด้วยข้อความ 5 ประเภทได้แก่ ข้อความเบรลล์ไทยล้วน ข้อความเบรลล์อังกฤษรูปเต็ม ข้อความเบรลล์อังกฤษรูปย่อ ข้อความเบรลล์ไทยปนอังกฤษรูปเต็ม และ ข้อความเบรลล์ไทยปนอังกฤษรูปย่อ ซึ่งได้มาจากศูนย์เทคโนโลยีการศึกษาเพื่อคนตาบอด

## 1.7 ขั้นตอนการวิจัย

1.7.1 เก็บรวบรวมข้อมูลที่ใช้ในการวิจัย

1.7.2 วิเคราะห์ระบบอักษรเบรลล์ไทย และระบบอักษรเบรลล์อังกฤษ

1.7.3 ดำเนินการวิจัยตามหัวข้อการออกแบบและพัฒนาระบบการถ่ายถอดอักษรเบรลล์ไทยปนอังกฤษ ซึ่งมีขั้นตอนดังนี้

1.7.3.1 รวบรวมข้อมูลที่ใช้ในการพัฒนาระบบ ได้แก่ รายการตัวย่อและคำย่อเบรลล์อังกฤษ ตัวบทเบรลล์ไทย ตัวบทเบรลล์อังกฤษ และตัวบทเบรลล์ไทยปนอังกฤษ

1.7.3.2 จัดทำคลังข้อความชุดฝึกสอนระบบ และดำเนินการฝึกสอนระบบโดยใช้ข้อมูลจากคลังข้อความภาษาไทยและคลังข้อความภาษาอังกฤษ

1.7.3.3 จัดทำชุดข้อความชุดทดสอบระบบ ได้แก่ กลุ่มที่ 1 ข้อความเบรลล์ไทย กลุ่มที่ 2 ข้อความเบรลล์อังกฤษ และกลุ่มที่ 3 ข้อความเบรลล์ไทยปนอังกฤษ

1.7.4 ออกแบบและพัฒนาระบบการถ่ายถอดอักษร เบรลล์ไทย เบรลล์อังกฤษ และเบรลล์ไทยปนอังกฤษ โดยแต่ละระบบจะใช้วิธีการ 3 รูปแบบคือ

1.7.4.1 ระบบการถ่ายถอดอักษร โดยใช้กฎทางภาษา

1.7.4.2 ระบบการถ่ายถอดอักษร โดยใช้แบบจำลองเอ็นแกรม

1.7.4.3 ระบบการถ่ายถอดอักษร โดยใช้กฎทางภาษาร่วมกับแบบจำลองเอ็นแกรม

1.7.5 ทดสอบการทำงานของแต่ละระบบ (ข้อ 1.7.4.1 – 1.7.4.3)

1.7.6 เปรียบเทียบประสิทธิภาพด้านความถูกต้องของแต่ละวิธีการ

1.7.7 วิเคราะห์ผลการวิจัย และสรุปผลการวิจัย

## 1.8 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบระบบของอักษรเบรลล์ไทยและระบบอักษรเบรลล์อังกฤษ

2. เป็นแนวทางในการศึกษาเปรียบเทียบระบบอักษรเบรลล์ในภาษาอื่น ๆ

3. เป็นแนวทางในการพัฒนาระบบการถ่ายถอดอักษรเบรลล์ในภาษาอื่น ๆ

## 1.9 โครงสร้างวิทยานิพนธ์

ในวิทยานิพนธ์ฉบับนี้จะแบ่งเนื้อหาในการนำเสนอออกเป็น 5 ส่วน ได้แก่

ในบทที่ 2 จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ได้แก่ ความรู้ทั่วไปเกี่ยวกับอักษรเบรลล์ การจำแนกภาษา การถ่ายถอดอักษร และงานวิจัยที่เกี่ยวข้องกับการถ่ายถอดเบรลล์ไทยปนอังกฤษ

ในบทที่ 3 จะกล่าวถึงการวิเคราะห์ระบบอักษรเบรลล์ไทยโดยการเปรียบเทียบตัวอักษรเบรลล์ไทยกับเบรลล์อังกฤษและเบรลล์ญี่ปุ่น และการวิเคราะห์อักษรวิธีของเบรลล์ไทยตามหลักภาษาศาสตร์

ในบทที่ 4 จะกล่าวถึง แนวคิดในการออกแบบระบบการถ่ายทอดอักษร ได้แก่ แนวคิดในการถ่ายทอดเบรลล์ไทย แนวคิดในการถ่ายทอดเบรลล์อังกฤษ และแนวคิดในการระบุภาษา

ในบทที่ 5 – บทที่ 7 จะได้นำเสนอ ระบบการถ่ายทอดเบรลล์ไทย ระบบการถ่ายทอดเบรลล์อังกฤษ และระบบการถ่ายทอดเบรลล์ไทยปนอังกฤษ โดยบทที่ 5 เสนอระบบการถ่ายทอดอักษรโดยใช้กฎ บทที่ 6 เสนอระบบการถ่ายทอดอักษรโดยใช้แบบจำลองเอ็นแกรม และบทที่ 7 เสนอระบบการถ่ายทอดอักษรโดยวิธีการแบบผสม

ในบทที่ 8 จะสรุปผลที่ได้จากวิทยานิพนธ์นี้ รวมทั้งให้ข้อเสนอแนะในการพัฒนาระบบการถ่ายทอดอักษรเบรลล์ให้เป็นอักษรปกติ ต่อไป

#### 1.10 รายการนิยามศัพท์ที่เกี่ยวข้อง

1. ระบบการถ่ายทอดอักษร (Transliteration system) หมายถึง กระบวนการในการเปลี่ยนข้อความอิเล็กทรอนิกส์จากอักษรเบรลล์เป็นอักษรปกติ โดยอาศัยคอมพิวเตอร์

2. อักษรเบรลล์ (Braille systems) หมายถึง ระบบตัวเขียนที่คนตาบอดใช้สื่อสาร โดยรับรู้ผ่านการสัมผัสด้วยปลายนิ้วมือ ประกอบด้วยเซลล์ต่าง ๆ นำมาเรียงต่อกันจากซ้ายไปขวา โดยที่แต่ละเซลล์ประกอบด้วยจุดจูนไม่เกิน 6 จุด การปรากฏของจุดในตำแหน่งที่ต่างกันใช้แทนอักขระต่างกัน

3. อักษรเบรลล์ไทยปนอังกฤษ (English-Thai braille) หมายถึง ข้อความอักษรเบรลล์ที่มีการสลับกันของภาษาไทยและภาษาอังกฤษ ในข้อความเดียวกัน

4. อักษรปกติ (Print systems) หมายถึง ระบบตัวเขียนสำหรับคนสายตาปกติที่ประกอบด้วย ตัวอักษร (letters) ตัวเลข (numbers) หรือสัญลักษณ์ (signs) ที่ปรากฏบนวัสดุที่มีพื้นผิวเรียบ เช่น กระดาษ หน้าจอแสดงผลของอุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ เป็นต้น

5. อักษรเบรลล์รูปเต็ม หรืออักษรเบรลล์เกรด 1 (Braille grade 1) หมายถึง อักษรเบรลล์หนึ่งหรือสองตัว ที่ใช้แทนตัวอักษรปกติหนึ่งตัว

6. อักษรเบรลล์รูปย่อ หรืออักษรเบรลล์เกรด 2 (Braille grade 2) หมายถึง อักษรเบรลล์หนึ่งตัวหรือมากกว่าหนึ่งตัวที่ใช้แทนตัวอักษรปกติหนึ่งกลุ่ม

7. คอมพิวเตอร์เบรลล์ (Computer braille) หมายถึง อักษรเบรลล์แต่ละตัวที่ใช้แทนอักขระแอสกีโดยตรงแบบหนึ่งต่อหนึ่ง

8. สัญลักษณ์แอสกี (ASCII symbols) หมายถึง สัญลักษณ์ที่ปรากฏบนหน้าจอคอมพิวเตอร์ ซึ่งใช้แทนอักษรเบรลล์ เช่น ‘g’ ในคอมพิวเตอร์จะใช้แทน ‘⠠’ และ ‘(’ ในคอมพิวเตอร์ใช้แทน ‘⠠’

9. เครื่องหมายนำเลข (Number sign; #) หมายถึง เครื่องหมายที่ใช้นำหน้าตัวอักษร 10 ตัวแรกในภาษาอังกฤษ (a-j) เพื่อบ่งชี้ว่าตัวอักษรที่ปรากฏหลังเครื่องหมายนี้เป็นตัวเลข ไม่ใช่ตัวอักษร

10. เครื่องหมายประกอบคำ (Composition sign) หมายถึง สัญลักษณ์ที่ปรากฏเฉพาะในอักษรเบรลล์อังกฤษเพื่อกำกับประเภทของอักษร ประกอบด้วย ‘⠨’ (เครื่องหมายแสดงตัวพิมพ์ใหญ่เดี่ยว) เป็นเครื่องหมายที่ใช้ระบุว่าอักษรตัวแรกที่ตามมาเป็นอักษรตัวพิมพ์ใหญ่ ‘⠨⠨’ (เครื่องหมายแสดงตัวพิมพ์ใหญ่คู่) ใช้ระบุว่าอักษรทั้งหมดที่ตามมาเป็นอักษรตัวพิมพ์ใหญ่ และ ‘⠨’ (เครื่องหมายกำกับตัวอักษร) เป็นเครื่องหมายที่บ่งชี้ว่าอักษรที่ตามมาเป็นตัวอักษรไม่ใช่ตัวย่อ

11. การจำแนกภาษา หรือการระบุภาษา (Language identification) หมายถึง การบอกว่าสายอักขระใดสายอักขระหนึ่งในอักษรเบรลล์นั้นต้องตีความให้เป็นภาษาใด โดยใช้คอมพิวเตอร์

12. วิธีการใช้กฎ (Rule-based method) หมายถึง การนำกฎทางภาษามาใช้ในการถอดออกอักษรเบรลล์ให้เป็นอักษรปกติ ซึ่งกฎที่ใช้ประกอบด้วย กฎในการจับคู่หน่วยอักขระระหว่างภาษา และกฎในการจัดลำดับของหน่วยอักขระในอักษรปกติ

13. วิธีการใช้แบบจำลองเอ็นแกรม (N-gram method) หมายถึง การถอดออกอักษรเบรลล์ให้เป็นอักษรปกติโดยอาศัยข้อมูลทางสถิติ โดยที่แบบจำลองเอ็นแกรมจะเป็นตัวเลือกรูปแบบที่จะนำมาใช้เป็นผลลัพธ์ของการถอดออกอักษร

14. วิธีการแบบผสม (Hybrid method) หมายถึง การถอดออกอักษรเบรลล์ให้เป็นอักษรปกติโดยการใช้กฎร่วมกับแบบจำลองเอ็นแกรม

15. คู่เทียบ หมายถึง อักษรเบรลล์และอักษรปกติที่ตรงกัน ซึ่งระบบจะใช้ในการถอดออกจากอักษรเบรลล์ให้เป็นอักษรปกติ

16. สายอักขระ หมายถึง ข้อความเบรลล์แต่ละวรรคที่นำมาถอดออกเป็นอักษรปกติ

17. เซ็ตของสายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา หมายถึง เซ็ตของอักขระเบรลล์ 1 ตัว 2 ตัว หรือ 3 ตัว ในภาษาหนึ่งที่แตกต่างกันจากอีกภาษาหนึ่ง ซึ่งระบบนำมาใช้ในการระบุภาษา

18. เซ็ตไตรแกรมท้ายพยางค์ หมายถึง เซ็ตของอักขระเบรลล์ 3 ตัว ที่ระบบนำมาใช้ในการตัดสินใจ กรณีที่พบอักขระเบรลล์ที่ตีความได้มากกว่าหนึ่งกรณี

19. สระพิเศษ หมายถึง สระที่จับคู่กับอักษรปกติได้แบบตัวต่อตัว แต่เมื่อสระเหล่านี้ปรากฏกับวรรณยุกต์จะมีการจัดเรียงที่แตกต่างจากอักษรปกติ ได้แก่ สระ 2 ตัวคือ ‘⠠⠨’ (ะ) และ ‘⠠⠨⠨’ (เ้า)

20. สระหน้า หมายถึง สระที่ปรากฏหน้าพยัญชนะต้นหรือปรากฏร่วมกับ ‘⠠⠨’ (ะ) ได้ ในเบรลล์ไทย ได้แก่ สระ 3 ตัว คือ ‘⠠⠨⠨’ (เ) ‘⠠⠨⠨⠨’ (แ) และ ‘⠠⠨⠨⠨⠨’ (โ)

## บทที่ 2

### ทบทวนวรรณกรรม

ในการวิจัยครั้งนี้ ผู้วิจัยจะนำเสนอเนื้อหาที่ได้ทำการศึกษาโดยแบ่งออกเป็นหัวข้อต่าง ๆ ได้แก่ หัวข้อที่ 2.1 ความรู้ทั่วไปเกี่ยวกับอักษรเบรลล์ เพื่อให้ผู้อ่านมีความเข้าใจพื้นฐานของการใช้อักษรเบรลล์ในการแทนตัวอักษรภาษา หัวข้อที่ 2.2 จะทบทวนงานด้านการจำแนกภาษาคำด้วยคอมพิวเตอร์ เนื่องจากข้อความเบรลล์ที่ใช้จริงมีการปนระหว่างภาษาซึ่งการจำแนกให้รู้ว่าเป็นเบรลล์ไทยหรืออังกฤษถือเป็นขั้นตอนแรกของการถอดถอดอักษร จากนั้นหัวข้อที่ 2.3 จะทบทวนงานด้านการถอดถอดอักษรระหว่างภาษาคำด้วยคอมพิวเตอร์ ในหัวข้อที่ 2.4 จะได้ทบทวนงานวิจัยที่เกี่ยวข้องกับการถอดถอดเบรลล์ไทยปนอังกฤษเป็นอักษรปกติ ที่ได้มีการศึกษาวิจัยมาก่อนหน้านี้ และในหัวข้อที่ 2.5 จะได้ทบทวนข้อมูลพื้นฐานเกี่ยวกับ โปรแกรม Duxbury Braille Translator เนื่องจากผู้วิจัยจะนำโปรแกรมนี้มาใช้เป็นเครื่องมือในการแปลงข้อความจากเบรลล์อังกฤษรูปย่อเป็นเบรลล์อังกฤษรูปเต็มเพื่อใช้ทดสอบระบบ

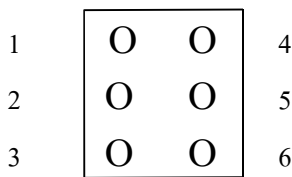
#### 2.1 ความรู้ทั่วไปเกี่ยวกับอักษรเบรลล์

ในการรับรู้ข้อมูลจากสิ่งพิมพ์ต่าง ๆ นั้น ผู้ที่มีสายตาปกติมักจะรับรู้ผ่านตัวอักษร เนื่องจากตัวอักษรเหล่านี้มักจะเป็นหมึกพิมพ์อยู่บนหน้ากระดาษ ทำให้ไม่สามารถรับรู้ด้วยประสาทสัมผัสอื่น ๆ ได้นอกจากสายตาเท่านั้น จึงจำเป็นที่จะต้องมียระบบอักษรเบรลล์สำหรับคนตาบอด ซึ่งเป็นระบบที่ประกอบด้วยเซลล์ต่าง ๆ นำมาเรียงต่อกัน แต่ละเซลล์ประกอบด้วยจุดขนาดไม่เกิน 6 จุด โดยที่การปรากฏของจุดในตำแหน่งที่ต่างกันใช้แสดงอักขระที่ต่างกัน ทำให้คนตาบอดสามารถรับรู้ข้อมูลภาษาได้โดยการสัมผัสด้วยปลายนิ้วมือ

อักษรเบรลล์แต่ละตัวหรือเรียกว่าแต่ละเซลล์ เป็นอักษรที่ใช้เนื้อที่ในการปรากฏเท่า ๆ กัน คือมีขนาดเนื้อที่พอที่จุดขนาดทั้ง 6 จุด จะสามารถเรียงต่อกันแบบ 3 แถว 2 คอลัมน์ได้<sup>1</sup> เพื่อให้เกิดความสะดวกในการอ้างถึงจุดต่าง ๆ จึงมีการเรียกชื่อจุดแต่ละจุด โดยเริ่มจากคอลัมน์ทางซ้ายจากบนลงล่างเป็นจุด 1, 2, 3 และ คอลัมน์ทางขวาจากบนลงล่างเป็นจุด 4, 5, 6 ตามลำดับ (Petzold, 2000: 17) ดังปรากฏในภาพที่ 2.1

<sup>1</sup> อักษรเบรลล์อังกฤษประกอบด้วยระบบ 6 จุด และ 8 จุด ที่นิยมใช้คือระบบ 6 จุด ส่วนระบบ 8 จุด จะนำมาใช้ในอุปกรณ์คอมพิวเตอร์แบบเบรลล์ (Sullivan, 2014:7) งานวิจัยชิ้นนี้จะศึกษาเฉพาะอักษรเบรลล์ระบบ 6 จุด เท่านั้น





ภาพที่ 2.1 การปรากฏของจุดในอักษรเบรลล์

เนื่องจากอักษรเบรลล์เกิดจากการเรียงต่อกันของจุดไม่เกิน 6 จุด จึงใช้แทนตัวอักษรได้เพียง 64 ตัวเท่านั้น แต่ในความเป็นจริงแล้วภาษาในโลกเป็นจำนวนมากมีหน่วยอักขระ (grapheme) มากกว่า 64 ตัว ทำให้อักษรเบรลล์ของแต่ละภาษามีโครงสร้างซับซ้อน เนื่องจากวิทยานิพนธ์ฉบับนี้มุ่งศึกษาเฉพาะภาษาอังกฤษและภาษาไทยเท่านั้น จึงจะกล่าวถึงการใช้อักษรเบรลล์สำหรับภาษาอังกฤษและภาษาไทยดังนี้

### 2.1.1 อักษรเบรลล์อังกฤษ

อักษรเบรลล์อังกฤษมี 3 รูปแบบคือ เบรลล์อังกฤษเกรด 1 เบรลล์อังกฤษเกรด 2 และเบรลล์อังกฤษเกรด 3 สำหรับเบรลล์อังกฤษเกรด 1 ซึ่งวิทยานิพนธ์ฉบับนี้จะเรียกว่าเบรลล์อังกฤษรูปเต็ม หมายถึงอักษรเบรลล์หนึ่งเซลล์ใช้แทนอักษรปกติหนึ่งตัว เบรลล์อังกฤษเกรด 2 ซึ่งวิทยานิพนธ์ฉบับนี้จะเรียกว่าเบรลล์อังกฤษรูปย่อ หมายถึงรหัสอักษรเบรลล์ที่ใช้แทนคำหรือบางส่วนของคำ และเบรลล์อังกฤษเกรด 3 หมายถึงอักษรเบรลล์รูปย่อที่ไม่ใช่รูปมาตรฐาน (non-standardized personal shorthand) ซึ่งใช้ในการจดบันทึกส่วนตัวเฉพาะบุคคลเท่านั้น (Klenk, 2011: 3-4) ในที่นี้จะขอกกล่าวถึงเฉพาะเบรลล์อังกฤษรูปเต็ม (Uncontracted English Braille) และเบรลล์อังกฤษรูปย่อ (Contracted English Braille) หรือเบรลล์เกรด 1 และเบรลล์เกรด 2 เท่านั้น

#### 2.1.1.1 อักษรเบรลล์อังกฤษรูปเต็ม

ในอักษรเบรลล์อังกฤษรูปแบบนี้ อักษรเบรลล์มีจำนวน 26 ตัว เท่ากับอักษรปกติ ดังปรากฏในตารางต่อไปนี้ (Petzold, 2000: 18)

ตารางที่ 2.1 ตัวอักษรเบรลล์อังกฤษ

แถว	ตัวอักษรเบรลล์อังกฤษ									
1	⠁	⠃	⠉	⠇	⠑	⠋	⠎	⠏	⠗	⠏
	a	b	c	d	e	f	g	h	i	j
2	⠅	⠇	⠍	⠏	⠕	⠎	⠏	⠗	⠑	⠗
	k	l	m	n	o	p	q	r	s	t
3	⠠	⠡	⠢	⠣	⠤					
	u	v	x	y	z					
4										⠠
										w

ตารางข้างต้นแสดงให้เห็นว่า เบรลล์อังกฤษเกิดจากการนำตัวอักษรหลัก 10 ตัว มาใช้เป็นฐานในการสร้างตัวอักษรอื่น ๆ ตัวอักษรหลักได้แก่ ‘⠁’ - ‘⠐’ (=a-j) ดังแสดงในแถวที่ 1 โดยใช้จุด 1, 2, 4, และ 5 ในการแยกความแตกต่างของตัวอักษร ส่วน ‘⠠’ - ‘⠤’ (=k-t) เกิดจากการใช้จุด 1, 2, 4, 5 ในลักษณะเดียวกับ a-j และเติม ‘⠠’ (จุด 3) *เข้าไปในเซลล์* ดังแสดงในแถวที่ 2 ส่วน ‘⠠’ - ‘⠤’ (=u-z (ยกเว้น w)) เกิดจากการนำตัวอักษรหลัก 5 ตัวแรก คือ a - e มาเติม ‘⠠’ (จุด 3-6) ดังแสดงในแถวที่ 3 และตัวอักษร ‘⠠’ (=w) เกิดจากการเติมเฉพาะ ‘⠠’ (จุด 6) ที่ตัว ‘⠐’ (=j) (Englebretson, 2009: 3; Petzold, 2000: 16)

ที่กล่าวมานี้เป็นเพียงอักษรเบรลล์อังกฤษตัวพิมพ์เล็กเท่านั้น หากเป็นตัวพิมพ์ใหญ่จะใส่เครื่องหมาย ‘⠠’ (capital sign) นำหน้าตัวอักษรตัวนั้น แต่หากเป็นตัวพิมพ์ใหญ่ทั้งคำก็ให้ใส่เครื่องหมาย ‘⠠⠠’ (double capital sign) นำหน้าคำ

การประสมอักษรในเบรลล์อังกฤษมีลักษณะการเขียนที่เรียงกันไปเหมือนตัวอักษรอังกฤษทั่วไปแบบหนึ่งต่อหนึ่ง โดยแต่ละคำจะมีการเขียนเว้นวรรค (หรือพิมพ์โดยใช้เครื่องหมาย space) ทำให้สามารถแยกคำต่าง ๆ ออกจากกันได้ เช่น "The dog" และ "THE DOG" สามารถเขียนได้ดังนี้

⠠⠠⠠⠠⠠⠠⠠⠠

T h e d o g

⠠⠠⠠⠠⠠⠠⠠⠠

T H E D O G

### 2.1.1.2 อักษรเบรลล์อังกฤษรูปย่อ

เนื่องจากอักษรเบรลล์เป็นอักษรที่มีขนาดใหญ่ต้องใช้เนื้อที่มากในการปรากฏบนกระดาษ ทำให้การอ่านและการเขียนอักษรระบบนี้ต้องสิ้นเปลืองทรัพยากรและใช้เวลามาก ด้วยข้อจำกัดดังกล่าวจึงมีการประดิษฐ์ตัวย่อเบรลล์เพื่อให้คนตาบอดใช้อักษรเบรลล์ได้สะดวกรวดเร็วยิ่งขึ้น (Tennessee council of the blind, 2008: 3)

ตัวย่อเบรลล์อังกฤษ คือการนำอักษรเบรลล์จำนวนหนึ่งมาใช้แทนคำหรือส่วนของคำภาษาอังกฤษโดยไม่ต้องเขียนทุกตัวอักษรของคำนั้น ซึ่ง Braille Authority of North America (2002: 5) ได้จำแนกตัวย่อเบรลล์อังกฤษออกเป็น 3 กลุ่ม ได้แก่ ตัวย่อเบรลล์เซลล์เดี่ยว (one-cell contractions) ตัวย่อเบรลล์เซลล์คู่ (two-cell contractions) และคำย่อเบรลล์ (short-form words)

#### 2.1.1.2.1 ตัวย่อเบรลล์เซลล์เดี่ยว

ตัวย่อเบรลล์เซลล์เดี่ยว คือตัวย่อที่อักษรเบรลล์หนึ่งเซลล์ใช้แทนคำหรือส่วนของคำได้ Braille Authority of North America (2002: 5-7) ได้แบ่งตัวย่อเบรลล์เซลล์เดี่ยวออกเป็นประเภทต่าง ๆ ได้แก่ ตัวย่ออักษรโดด ตัวย่อรูปเต็มคงรูป ตัวย่อรูปเต็มกลายรูป ตัวย่อประกอบคำ ตัวย่อกลางคำหรือท้ายคำ และตัวย่อเบรลล์ต่ำ

##### (1) ตัวย่ออักษรโดด (alphabet contractions)

ในเบรลล์อังกฤษ มีอักษรเบรลล์ใช้แทนตัวอักษร 26 ตัวเช่นเดียวกับอังกฤษทั่วไป และตัวอักษรเหล่านี้สามารถนำมาเรียงต่อกันเป็นคำได้เช่นเดียวกัน เช่น ‘:::’ สามารถเทียบเท่ากับอักษรปกติ “ant” แบบตัวต่อตัวได้ ทำให้อักษรเบรลล์เหล่านี้เมื่อมีอักษรอื่นปรากฏร่วมหน้าและ/หรือหลังภายในคำจะใช้แทนอักษรอังกฤษทั่วไป อย่างไรก็ตาม ในอังกฤษรูปย่ออักษรเบรลล์เหล่านี้สามารถนำมาใช้แทนคำโดดได้ กล่าวคือเมื่ออักษรแต่ละตัวในกลุ่มนี้ปรากฏระหว่างช่องว่างและ/หรือเครื่องหมายแสดงขอบเขตของคำ จะใช้แทนคำโดด เช่น ‘|::|’ ซึ่งเป็นอักษร b จะใช้แทน “|but|”<sup>2</sup> ‘|:::|’ ซึ่งเป็นอักษร r กับ ? จะใช้แทน “|from?” เป็นต้น เป็นที่น่าสังเกตว่าอักษรเหล่านี้ไม่สามารถใช้เป็นส่วนหนึ่งของคำได้ เช่น ‘::’ ใช้แทน “but” แต่คำว่า “butter” ไม่สามารถเขียนเป็น ‘:::’ (= bter) ได้

ตัวย่ออักษรโดดจะใช้เป็นส่วนหนึ่งของคำได้ในกรณีต่อไปนี้

○ ปรากฏกับ ‘::’ (hyphen) เช่น ‘:::’ (=m-er-r-y---go---r-ound) ซึ่งใช้แทน “merry-go-round” ที่มี ‘::’ (= go) ปรากฏระหว่างเครื่องหมาย hyphen

<sup>2</sup> ในงานวิจัยชิ้นนี้ จะใช้เครื่องหมาย | แทนเครื่องหมายเว้นวรรคในข้อความภาษาอังกฤษ

- ใช้กับคำที่มี ‘:’ (apostrophe) เช่น ‘::: :::’ (=c’t) เพื่อย่อ “can’t” เป็นต้น

กล่าวโดยสรุป ตัวย่ออักษรคำโดดจะใช้อักษรเบรลล์หนึ่งตัวแทนคำคงที่หนึ่งคำ ไม่สามารถปรากฏกับอักษรอื่นได้ ยกเว้นการใช้ติดกับอักษรที่แทนเครื่องหมายต่าง ๆ เช่น “-”, “:”, “ ‘ ”, “?” เป็นต้น

การใช้ตัวย่ออักษรโดดที่แสดงใน (Braille Authority of North America, 2002: XI-2) มีดังนี้

ตารางที่ 2.2 การใช้ตัวย่ออักษรโดด

วิธีการใช้	ตัวอย่าง		
	รูปย่อในเบรลล์	แอสกี	คำหรือประโยค
1) ใช้อักษรตามคำฟังเพื่อย่อคำโดด	⠠⠠⠠⠠⠠⠠⠠⠠	n y b p l y	Not you but people like you.
2) ใช้หน้าหรือหลังเครื่องหมาย hyphen ในกรณีที่เป็นคำประสม	⠠⠠⠠⠠⠠⠠⠠⠠	s-call\$	<u>so-called</u>
3) ใช้กับคำวิสามัญนาม	⠠⠠⠠⠠⠠⠠⠠⠠	,mr. ,m	Mr. <u>More</u>
4) ใช้กับคำแสดงความเป็นเจ้าของ	⠠⠠⠠⠠⠠⠠⠠⠠	,w’s	<u>Will's</u>
5) ใช้กับคำที่ค้นด้วย apostrophe	⠠⠠⠠⠠⠠⠠⠠⠠	c’t	<u>can't</u>

ตัวย่ออักษรโดดในกลุ่มเบรลล์อังกฤษ 26 ตัว มีรูปย่อและสิ่งที่ใช้แทนที่เป็นมาตรฐานสำหรับเบรลล์อังกฤษทั่วโลก (Braille Authority of North America, 2002: 2) ดังนี้

ตารางที่ 2.3 รายการตัวย่ออักษรโดด

เบรลล์	แอสกี	รูปเต็ม
⠠	b	but
⠠	c	can
⠠	d	do
⠠	e	every
⠠	f	from
⠠	g	go
⠠	h	have
⠠	j	just
⠠	k	knowledge
⠠	l	like
⠠	m	more
⠠	n	not
⠠	p	people
⠠	q	quite
⠠	r	rather
⠠	s	so
⠠	t	that
⠠	u	us
⠠	v	very
⠠	w	will
⠠	x	it
⠠	y	you
⠠	z	as

(หมายเหตุ: สัญลักษณ์แอสกีที่ปรากฏในตาราง หมายถึงสัญลักษณ์แอสกีที่ใช้แทนตัวย่อ โดยการป้อนข้อมูลผ่านคีย์บอร์ด)

เป็นที่น่าสังเกตว่าในการปรากฏของตัวอักษรโศด เบลล์แต่ละเซลล์จะใช้แทนคำแต่ละคำ ทำให้ไม่สามารถเขียนอักษรเบลล์เหล่านี้เพื่อหมายถึงอักษรแต่ละตัวตามลำพังได้ ฉะนั้นเมื่อต้องการแสดงตัวอักษรโศด ต้องใช้ ‘:’ (เครื่องหมายกำกับตัวอักษร) นำหน้าอักษรเบลล์เหล่านี้ เช่นเมื่อพบ ‘:’ (=b) ปรากฏระหว่างช่องว่างต้องตีความเป็นคำว่า “but” แต่หากหมายถึงตัวอักษร b เท่านั้นจำเป็นต้องใส่เครื่องหมายกำกับอักษรไว้หน้าตัว b ทำให้ได้เป็น ‘: :’ จึงจะตีความเป็นตัวอักษร b ได้

(2) ตัวอรรูปเต็มคกรรพ ( whole-word/part-word contractions type1)

ตัวอรรูปเต็มคกรรพ เป็นตัวอรรที่อักษรเบลล์ 1 ตัว ใช้แทนกรรพตัวอักษรที่ตายตัวไม่วาอักษรนั้นจะปรากฏเป็นคำโศดหรือเป็นส่วนของคำก็ตมยงคงใช้แทนกรรพตัวอักษรเดิม เช่น ‘|:|’ (= &) ซึ่งใช้แทนคำ “and” ที่ปรากฏเป็นคำโศด และ ‘: : :’ (= g-r-and) ซึ่งมี “and” เป็นส่วนหนึ่งของคำ

จะเห็นได้วา คำที่ได้จากตัวอรรในกรรพนี้มีควมถึในการปรากฏสูงในภษาอังกฤษ อีกรั้งคำเหล่านี้มักปรากฏเรียงต่อกัน จึงมีการเขียนตัวอรรเหล่านี้ต่อกันโดยไม่เว้นวรรคเพื่อใช้แทนคำโศดแต่ละคำ (ซึ่งคำเหล่านี้ไม่สามารถเขียนชิดกันโดยไม่เว้นวรรคได้ในอักษรปกติ) เช่น ‘: : :’ (= &! ) ใช้ย่อ “and the” ซึ่งเป็นคำโศด 2 คำที่เรียงต่อกัน และ ‘: : : :’ (= &a) ที่ใช้แทนคำโศด 3 คำที่อยู่ชิดกันคือ “and with a” เป็นต้น

ตัวอรรูปเต็มคกรรพ มีตัวอย่างการใช้ซึ่งแสดงได้ดังนี้ (Braille Authority of North America, 2002: XI-3)

ตารางที่ 2.4 วิธีการใช้ตัวอรรูปเต็มคกรรพ

วิธีการใช้	ตัวอย่าง		
	สัญลักษณ์แอสกี	รอรย่อในเบลล์	คำหรือประโยค
1) เมื่อใช้ตัวอรรกับคำโศดอื่น ๆ ให้เว้นวรรคระหว่างตัวอรรและคำโศด	,i l hm & h(er).	: : : : : : : :	I like him and her.
2) เมื่อตัวอรรอยู่ติดกันให้เขียนติดต่อกันไม่เว้นวรรค	! cat & ! rat	: : : : : : : : : : : :	the cat <u>and</u> the rat.
3) ใช้ย่อส่วนของคำ	! sis	: : : : :	<u>T</u> hesis

ตัวอรรูปเต็มคกรรพ ประกอบด้วยตัวอรรและกรรพตัวอักษรที่ใช้แทน ดังนี้

ตารางที่ 2.5 รายการตัวย่อรูปเต็มกรุงรูป

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม
⠆	&	and
⠆	=	for
⠆	(	of
⠆	!	the
⠆	)	with

## (3) ตัวย่อรูปเต็มกลายรูป (whole-word/part-word contractions type 2)

ตัวย่อรูปเต็มกลายรูป เป็นตัวย่อที่ใช้แทนกลุ่มตัวอักษรที่แตกต่างกันใน 2 บริบท คือ เมื่อปรากฏเป็นคำโดดจะใช้แทนตัวอักษรกลุ่มหนึ่ง และเมื่อปรากฏเป็นส่วนของคำก็จะใช้แทนตัวอักษรอีกกลุ่มหนึ่ง เช่น ‘⠆’ จะใช้แทน “which” เมื่อปรากฏเป็นคำโดด แต่เมื่อปรากฏภายในคำจะใช้แทน “wh” ตัวย่อกลุ่มนี้ไม่เป็นปัญหาต่อการจดจำเท่าใดนัก เนื่องจากจะใช้แทนกลุ่มตัวอักษรที่มีอักษร 2 ตัวแรกซ้ำกัน เช่น ‘⠆’ ใช้ย่อ “shall” หากเป็นคำโดด และใช้ย่อ “sh” เมื่อปรากฏเป็นส่วนหนึ่งของคำ ซึ่งอักษร 2 กลุ่มนี้มี “sh” ร่วมกัน

ตัวย่อรูปเต็มกลายรูป มีรูปย่อและสิ่งที่ใช้แทนดังนี้ (Braille Authority of North America, 2002: XI-4)

ตารางที่ 2.6 รายการตัวย่อรูปเต็มกลายรูป

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม	
		คำโดด	ส่วนของคำ
⠆	*	child	ch
⠆	\	out	ou
⠆	%	shall	sh
⠆	/	still	st
⠆	?	this	th
⠆	:	which	wh

## (4) ตัวย่อประกอบคำ (part-word contractions)

ตัวย่อประกอบคำ เป็นตัวย่อที่ต้องปรากฏเป็นส่วนหนึ่งของคำ ใช้สำหรับแทนส่วนของคำที่มีตัวอักษรเหล่านี้คือ “ar”, “ed”, “er”, “gh” และ “ow” ตัวย่อประกอบคำมีรูปย่อและสิ่งที่ใช้แทนดังนี้

ตารางที่ 2.7 รายการตัวย่อประกอบคำ

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม
⠠⠠	>	ar
⠠⠠⠠	\$	ed
⠠⠠⠠⠠	]	er
⠠⠠⠠⠠⠠	<	gh
⠠⠠⠠⠠⠠⠠	[	ow

ตัวย่อเหล่านี้มีวิธีการใช้ดังนี้ (Braille Authority of North America, 2002: XII-3)

ตารางที่ 2.8 วิธีการใช้ตัวย่อประกอบคำ

วิธีการใช้	ตัวอย่าง		
	คำหรือประโยค	รูปย่อ	แอสกี
1) ใช้ตัวย่อในที่ตำแหน่งต้น กลางหรือท้ายคำ	error	⠠⠠⠠⠠⠠	]ror
2) ใช้แสดงเสียงของคำ	ar! (หัวเราะ)	⠠⠠⠠	>6

## (5) ตัวย่อกลางคำหรือท้ายคำ (middle or end of word contractions)

ตัวย่อกลางคำหรือท้ายคำ หมายถึงตัวย่อที่ปรากฏในตำแหน่งกลางคำหรือท้ายคำได้เท่านั้น ตัวย่อกลุ่มนี้เป็นตัวย่อที่ใช้เบรลล์เซลล์เดียวเพื่อแทนกลุ่มตัวอักษร ประกอบด้วยตัวย่อ 2 ตัวได้แก่ ⠠⠠ (=ble) และ ⠠⠠ (=ing)

เป็นที่น่าสังเกตว่า อักษรเบรลล์ 2 ตัวที่กล่าวมาข้างต้นนี้เมื่อปรากฏต้นสายอักษรจะทำหน้าที่อื่น ๆ ที่ไม่ใช่ตัวย่อ ดังที่แสดงไว้ในหัวข้อที่เกี่ยวกับตัวเลขเบรลล์ (ดูหัวข้อ 2.1.3 ประกอบ) ตัวย่อกลางคำหรือท้ายคำ มีรูปย่อและสิ่งที่ใช้แทน ดังนี้



ตารางที่ 2.9 รายการตัวย่อกลางคำหรือท้ายคำ

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม
::	#	ble
::	+	ing

ตัวย่อกลางคำหรือท้ายคำมีวิธีการใช้ดังนี้ (Braille Authority of North America, 2002: XII-2)

ตารางที่ 2.10 วิธีการใช้ตัวย่อกลางคำหรือท้ายคำ

วิธีการใช้	ตัวอย่าง		
	คำหรือประโยค	รูปย่อในเบรลล์	สัญลักษณ์แอสกี
1) ใช้ที่ตำแหน่งกลางคำ	Singer	::::	s+] ]
2) ใช้ที่ตำแหน่งท้ายคำ	Trouble	:::::	tr#

#### (6) ตัวย่อเบรลล์ต่ำ (lower-sign contractions)

หมายถึงตัวย่อเซลล์เดี่ยวที่ไม่มีจุด 1 และ/หรือจุด 4 ปรากฏอยู่ ตัวย่อเบรลล์ต่ำจำแนกเป็น 3 ประเภทตามตำแหน่งที่ปรากฏในคำ ได้แก่ ตัวย่อเบรลล์ต่ำย่อคำโดด ตัวย่อเบรลล์ต่ำย่อคำโดดพิเศษ และตัวย่อเบรลล์ต่ำย่อส่วนของคำ

ก. ตัวย่อเบรลล์ต่ำย่อคำโดด (lower-cell word contractions for “be, enough, were, his, in, was”)

ตัวย่อเบรลล์ต่ำย่อคำโดดมีลักษณะคล้ายกับตัวย่ออักษรโดด กล่าวคือจะต้องปรากฏตามลำพัง (stand alone) เท่านั้น ตัวย่อเหล่านี้มีรูปย่อและสิ่งที่ใช้แทนดังนี้ (Braille Authority of North America, 2002: XIII-2)

<sup>3</sup> เหตุที่เรียกว่าเบรลล์ต่ำเนื่องจากอักษรเหล่านี้ปรากฏในตำแหน่งที่ต่ำกว่าอักษรเบรลล์กลุ่มอื่น ๆ แม้ว่าคู่มือจะมิได้มีรูปเบรลล์เหมือนกันต้องอาศัยบริบทข้างเคียงเพื่อบอกว่าเป็นเบรลล์ต่ำหรือไม่ เช่น อักษร :: จะมีรูปซ้ำกับอักษร :: (ซึ่งอักษรตัวหลังถือว่าเป็นอักษรเบรลล์เซลล์ต่ำ) สัญลักษณ์เบรลล์ต่ำเหล่านี้บางตัวเมื่อปรากฏในตำแหน่งต้นหรือท้ายคำ จะถูกตีความให้เป็นเครื่องหมายวรรคตอน (punctuation) เช่น :: ถ้าปรากฏหน้าคำก็จะหมายถึงวงเล็บเปิด [ ( ) แต่หากอยู่ท้ายคำจะถูกตีความให้เป็นเครื่องหมายวงเล็บปิด [ ) ] (ดูตารางที่ 2.13 ประกอบ)

ตารางที่ 2.11 รายการตัวย่อเบรลล์ต่ำย่อคำโดด

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม
⠆	2	be
⠆⠆	5	enough
⠆⠆	7	were
⠆⠆	8	his
⠆⠆	9	in
⠆⠆	0	was

ข. ตัวย่อเบรลล์ต่ำย่อคำ โดดพิเศษ (lower-cell word contractions for “to, into, by”)

ตัวย่อกลุ่มนี้ลักษณะคล้ายกับตัวย่อเบรลล์ต่ำย่อคำโดดในแง่ที่ว่า ต่างก็ใช้ย่อคำโดด แต่ในการใช้ตัวย่อประเภทหลังนี้จำเป็นต้องเขียนติดกับคำต่อไปเลยไม่ต้องมีช่องว่าง เช่น ‘⠆⠆⠆’ ใช้เขียน “to go” และ ‘⠆⠆⠆⠆⠆⠆⠆⠆⠆’ ใช้เขียน “into water” เป็นต้น

รูปย่อและสิ่งที่ใช้แทนตัวย่อเบรลล์ต่ำย่อคำโดดพิเศษมีดังนี้ (Braille Authority of North America, 2002: XIII-5)

ตารางที่ 2.12 รายการตัวย่อเบรลล์ต่ำที่ย่อคำโดดพิเศษ

เบรลล์	สัญลักษณ์แอสกี	รูปเต็ม
⠆⠆	6	to
⠆⠆⠆⠆	96	into
⠆⠆	0	by

ค. ตัวย่อเบรลล์ต่ำย่อส่วนของคำ (part-word lower-cell contractions)

ตัวย่อเหล่านี้จะต้องใช้ประกอบคำเสมอ การปรากฏของตัวย่อแต่ละตัวในตำแหน่งที่ต่างกัน คือ ต้นคำ กลางคำ หรือท้ายคำ จะทำให้ตีความได้แตกต่างกัน ตัวอย่างเช่น ‘⠆⠆⠆⠆⠆⠆⠆⠆⠆’ (=con-v-e-y) ซึ่งใช้เขียนคำว่า “convey” เป็นต้น

ตัวย่อในตำแหน่งต่าง ๆ และสิ่งที่ใช้แทน มีดังนี้ (Braille Authority of North America, 2002: XIII-7)

ตารางที่ 2.13 รายการตัวย่อเบรลล์ต่ำย่อส่วนของคำ

เบรลล์	สัญลักษณ์ แอสกี	รูปเต็ม		
		ต้นคำ	กลางคำ	ท้ายคำ
⠆	1		ea	,
⠇	2	be	bb	;
⠈	3	con	cc	:
⠉	4	dis	dd	.
⠊	5	en	en	en
⠋	6		ff	!
⠌	7	(	gg	)
⠍	8	“		?
⠎	9	in	in	in
⠏	0			”
⠒	-	com	-	-

#### 2.1.1.2.2 ตัวย่อเบรลล์เซลล์คู่ (two-cell contractions)

ตัวย่อเบรลล์เซลล์คู่หมายถึง เบรลล์ 2 เซลล์ ที่ใช้แทนคำโดดหรือส่วนของคำ โดยที่เบรลล์เซลล์แรกจะขึ้นต้นด้วย จุด 5, 4-6, 5-6 หรือจุด 4-5-6

ตัวย่อกลุ่มนี้มี 2 ประเภท (Braille Authority of North America, 2002: XIII-9) ได้แก่

(1) ตัวย่อหน่วยหลัก (initial-letter contractions) เป็นตัวย่อเซลล์คู่ที่ปรากฏตามลำพังและสามารถปรากฏเป็นส่วนหนึ่งของคำได้ ตัวย่อนี้จะขึ้นต้นด้วยจุด 5, 4-5 หรือ 4-5-6 ตัวย่อหน่วยหลักมีรูปย่อและสิ่งที่ใช้แทนดังนี้

ตารางที่ 2.14 รายการตัวย่อหน่วยหลัก

เบรลล์	แอสกี	รูปเต็ม	เบรลล์	แอสกี	รูปเต็ม
⠠⠨	_c	cannot	⠠⠨	"s	some
⠠⠨	"*	character	⠠⠨	_s	spirit
⠠⠨	"d	day	⠠⠨	_!	their
⠠⠨	"e	ever	⠠⠨	"!	there
⠠⠨	"f	father	⠠⠨	^!	these
⠠⠨	_h	had	⠠⠨	^?	those
⠠⠨	"h	here	⠠⠨	"?	through
⠠⠨	"k	know	⠠⠨	"t	time
⠠⠨	"l	lord	⠠⠨	"u	under
⠠⠨	_m	many	⠠⠨	^u	upon
⠠⠨	"m	mother	⠠⠨	":	where
⠠⠨	"o	one	⠠⠨	^:	whose
⠠⠨	"p	part	⠠⠨	^w	word
⠠⠨	"q	question	⠠⠨	"w	work
⠠⠨	"r	right	⠠⠨	_w	world
⠠⠨	"\	ought	⠠⠨	"y	young

ตัวย่อหน่วยหลักมีวิธีการใช้ดังแสดงในตารางต่อไปนี้ (Braille Authority of North America, 2002: XIII-9)

ตารางที่ 2.15 การใช้ตัวย่อหน่วยหลัก

วิธีการใช้	ตัวอย่าง		
	คำหรือประโยค	รูปย่อในเบรลล์	สัญลักษณ์แอสกี
1) ใช้ย่อทั้งคำ	Part	⠠⠠⠠	"p
2) ใช้ย่อส่วนของคำ	apartment	⠠⠠⠠⠠⠠⠠	a"p;t

จากรายการตัวย่อในตารางข้างต้น จะพบว่ามีการใช้ ‘⠠’ (จุด 5) เพื่อนำหน้าเบรลล์เซลล์คู่เป็นหลัก แต่เนื่องจากการใช้ ‘⠠’ เพียงอย่างเดียวทำให้ไม่สามารถจะย่อคำในภาษาอังกฤษได้ครบถ้วน ดังนั้นจึงมีการใช้ ‘⠠’ (จุด 4-5) และ ‘⠠’ (จุด 4-5-6) เพื่อให้สามารถย่อคำอื่น ๆ ได้ เช่น

⠠⠠⠠ (จุด 5 w) ย่อ “work” ⠠⠠⠠⠠ (จุด 4-5 w) ย่อ “word” และ ⠠⠠⠠⠠⠠ (จุด 4-5-6 w) ย่อ “world” ดังนั้นการวิจัยครั้งนี้จึงมุ่งที่จะวิเคราะห์ระบบตัวย่อเบรลล์อังกฤษเพื่อนำไปใช้ในการจัดการเรียนการสอนและการประมวลผลทางคอมพิวเตอร์ สำหรับผลการวิเคราะห์จะนำเสนอในหัวข้อที่เกี่ยวข้องต่อไป

(2) ตัวย่อหน่วยเสริม (final-letter contractions) หมายถึง ตัวย่อเบรลล์เซลล์คู่ที่สามารถปรากฏได้ในตำแหน่งกลางคำหรือท้ายคำเท่านั้น โดยเซลล์แรกของตัวย่อเหล่านี้ขึ้นต้นด้วย ⠠⠠ (จุด 6) ⠠⠠⠠ (จุด 5-6) หรือ ⠠⠠⠠⠠ (จุด 4-6)

ตัวย่อหน่วยเสริมและสิ่งที่ใช้แทน แสดงได้ดังนี้ (Braille Authority of North America, 2002: XIII-9)

ตารางที่ 2.16 รายการตัวย่อหน่วยเสริม

เบรลล์	แอสกี	รูปเต็ม	เบรลล์	แอสกี	รูปเต็ม
⠠⠠⠠	,l	ful	⠠⠠⠠⠠	.t	ount
⠠⠠⠠⠠	,n	ation	⠠⠠⠠⠠⠠	;e	ence
⠠⠠⠠⠠	,y	ally	⠠⠠⠠⠠⠠	;g	ong
⠠⠠⠠⠠	.d	ound	⠠⠠⠠⠠⠠	;n	tion
⠠⠠⠠⠠	.e	ance	⠠⠠⠠⠠⠠	;s	ness
⠠⠠⠠⠠⠠	.n	sion	⠠⠠⠠⠠⠠	;t	ment
⠠⠠⠠⠠	.s	less	⠠⠠⠠⠠⠠	;y	ity

ตัวย่อหน่วยเสริมมีวิธีการใช้คือให้ใช้ที่ตำแหน่งกลางคำหรือท้ายคำ เช่น ⠠⠠⠠⠠⠠ (=d-ance) ใช้แทนคำว่า “dance” และ ⠠⠠⠠⠠⠠⠠ (=d-ance-r) ใช้แทน “dancer” เป็นต้น หากในการย่อคำ หนึ่งสามารถใช้ตัวย่อได้ 2 กรณี คือ ใช้ตัวย่อหน่วยเสริมเป็นส่วนประกอบของคำได้หรือใช้ตัวย่ออื่นแทนได้ ให้ใช้ตัวย่อหน่วยเสริมแทนที่จะใช้ตัวย่อในกลุ่มอื่น เช่นคำว่า “dancer” ต้องใช้ ⠠⠠⠠⠠⠠⠠ (=d-ance-r) ไม่สามารถเขียนเป็น ⠠⠠⠠⠠⠠⠠⠠ (=d-a-n-c-er) ได้

จากตารางที่ 2.16 จะเห็นว่า ตัวย่อกลุ่มนี้ต้องขึ้นต้นด้วย ⠠⠠ (จุด 6) ⠠⠠⠠ (จุด 4-6) หรือ ⠠⠠⠠⠠ (จุด 5-6) ตัวย่อประเภทนี้ต้องปรากฏเฉพาะตำแหน่งกลางคำหรือท้ายคำได้เท่านั้น เนื่องจากสัญลักษณ์เหล่านี้ หากปรากฏที่ตำแหน่งต้นคำจะใช้แทนเครื่องหมายอื่นที่ไม่ใช่ตัวย่อเบรลล์ ดังนี้

- ⠠⠠ (จุด 6) แสดงว่าตัวอักษรที่ตามมาตัวแรกเป็นอักษรตัวใหญ่ (capital sign)

- ‘:’ (จุด 4-6) แสดงว่า ตัวอักษรหลังเครื่องหมายนี้จนถึงเว้นวรรคมีกรณัน (เช่นเป็นอักษรตัวเอียง ตัวหนา หรือขีดเส้นใต้) ที่คำนั้น
- ‘:’ (จุด 5-6) แสดงว่า ตัวอักษรหลังเครื่องหมายนี้จนถึงเว้นวรรคเป็นตัวอักษร ไม่ใช่ตัวย่อหรือคำย่อ

#### 2.1.1.2.3 คำย่อเบรลล์ (shot-form words)

คำย่อเบรลล์หมายถึง คำในอักษรเบรลล์ที่เกิดจากการนำเอาอักษรบางตัวภายในคำมาใช้แทนคำนั้น ๆ เช่น ‘: : : : :’ (=rcvrg) ใช้แทน receiving

คำย่อเบรลล์ที่ปรากฏในภาษาอังกฤษทั้งหมด มีรูปย่อและสิ่งที่ใช้แทนดังนี้

## ตารางที่ 2.17 รายการคำย่อเบรลล์

เบรลล์	แอสกี	รูปเต็ม
⠠⠍⠋⠧⠎	<i>!mvs</i>	themselves
⠠⠋⠋⠋	<i>"of</i>	oneself
⠠⠎⠋⠇	<i>%d</i>	should
⠠⠎⠋⠗	<i>*n</i>	children
⠠⠎⠋⠋⠋	<i>?yf</i>	thysself
⠠⠋⠋⠋⠎	<i>\rvs</i>	ourselves
⠠⠎⠋	<i>2c</i>	because
⠠⠎⠋	<i>2f</i>	before
⠠⠎⠋	<i>2h</i>	behind
⠠⠎⠋	<i>2l</i>	below
⠠⠎⠋	<i>2n</i>	beneath
⠠⠎⠋	<i>2s</i>	beside
⠠⠎⠋	<i>2t</i>	between
⠠⠎⠋	<i>2y</i>	beyond
⠠⠎⠋⠎	<i>3cv</i>	conceive
⠠⠎⠋⠎⠎	<i>3cvg</i>	conceiving
⠠⠎⠋	<i>ab</i>	about
⠠⠎⠋⠎	<i>abv</i>	above
⠠⠎⠋	<i>ac</i>	according
⠠⠎⠋⠎	<i>acr</i>	across
⠠⠎⠋	<i>af</i>	after
⠠⠎⠋⠎	<i>afn</i>	afternoon
⠠⠎⠋⠎	<i>afw</i>	afterward
⠠⠎⠋	<i>ag</i>	again
⠠⠎⠋⠎	<i>ag/</i>	against
⠠⠎⠋	<i>al</i>	also
⠠⠎⠋⠎	<i>al?</i>	although
⠠⠎⠋⠎	<i>alm</i>	almost
⠠⠎⠋⠎	<i>alr</i>	already
⠠⠎⠋⠎	<i>alt</i>	altogether
⠠⠎⠋⠎	<i>alw</i>	always
⠠⠎⠋	<i>bl</i>	blind

เบรลล์	แอสกี	รูปเต็ม
⠠⠎⠋	<i>brl</i>	braille
⠠⠎⠋	<i>cd</i>	could
⠠⠎⠋⠎	<i>dcl</i>	declare
⠠⠎⠋⠎⠎	<i>dclg</i>	Declaring
⠠⠎⠋⠎	<i>dcv</i>	deceive
⠠⠎⠋⠎⠎	<i>dcvg</i>	deceiving
⠠⠎⠋	<i>ei</i>	either
⠠⠎⠋	<i>f/</i>	first
⠠⠎⠋	<i>fr</i>	friend
⠠⠎⠋	<i>Gd</i>	good
⠠⠎⠋⠎	<i>grt</i>	great
⠠⠎⠋⠎	<i>hjf</i>	herself
⠠⠎⠋	<i>hm</i>	him
⠠⠎⠋⠎	<i>hmf</i>	himself
⠠⠎⠋⠎	<i>imm</i>	immediate
⠠⠎⠋	<i>ll</i>	little
⠠⠎⠋	<i>lr</i>	letter
⠠⠎⠋	<i>m*</i>	much
⠠⠎⠋	<i>m/</i>	must
⠠⠎⠋⠎	<i>myf</i>	myself
⠠⠎⠋⠎	<i>nec</i>	necessary
⠠⠎⠋⠎	<i>nei</i>	neither
⠠⠎⠋⠎	<i>o'clock</i>	o'clock
⠠⠎⠋⠎⠎	<i>p/cv</i>	perceive
⠠⠎⠋⠎⠎⠎	<i>p/cvg</i>	perceiving
⠠⠎⠋⠎	<i>p/h</i>	perhaps
⠠⠎⠋	<i>pd</i>	paid
⠠⠎⠋	<i>qk</i>	quick
⠠⠎⠋⠎	<i>rcv</i>	receive
⠠⠎⠋⠎⠎	<i>rcvg</i>	receiving
⠠⠎⠋⠎	<i>rjc</i>	rejoice
⠠⠎⠋⠎⠎	<i>rjcg</i>	rejoicing

เบรลล์	แอสกี	รูปเต็ม
⠠⠠⠠	<i>s*</i>	such
⠠⠠⠠⠠	<i>sd</i>	said
⠠⠠⠠⠠	<i>td</i>	today
⠠⠠⠠⠠⠠	<i>tgr</i>	together
⠠⠠⠠⠠	<i>tm</i>	tomorrow
⠠⠠⠠⠠	<i>tn</i>	tonight

เบรลล์	แอสกี	รูปเต็ม
⠠⠠⠠⠠	<i>wd</i>	would
⠠⠠⠠⠠	<i>xf</i>	itself
⠠⠠⠠⠠	<i>xs</i>	its
⠠⠠⠠⠠	<i>yr</i>	your
⠠⠠⠠⠠⠠	<i>yrf</i>	yourself
⠠⠠⠠⠠⠠⠠	<i>yrvs</i>	yourselves

คำย่อเบรลล์ มีวิธีการใช้ดังนี้ (Braille Authority of North America, 2002: XIII-10)

#### ตารางที่ 2.18 การใช้คำย่ออักษรเบรลล์

วิธีการใช้	ตัวอย่าง		
	คำหรือประโยค	รูปย่อในเบรลล์	แอสกี
1) ใช้ย่อกำโดด	He is my <u>friend</u> .	⠠⠠⠠⠠ ⠠⠠⠠⠠ ⠠⠠⠠⠠	,he is my fr4
2) ใช้ย่อส่วนของคำ	<u>accordingly</u>	⠠⠠⠠⠠⠠	acly
3) ใช้กับคำวิสามานยนามที่เป็นคำโดด	Dr. <u>Good</u>	⠠⠠⠠⠠⠠ ⠠⠠⠠⠠	,dr4 ,gd
4) หากคำเต็มใช้รูปเดียวกับคำย่อให้ใส่ ‘⠠⠠’ (เครื่องหมายกำกับตัวอักษร) ที่หน้าคำ	<u>Al</u> Johnson	⠠⠠⠠⠠⠠ ⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠	,;al ,johnson
5) กรณีคำย่อเป็นส่วนหนึ่งของคำ หากแยกคำออกจากกันให้รักษารูปคำย่อไว้	mis- <u>conceived</u>	⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠	mis- 3cvd

เป็นที่น่าสังเกตว่าในการปรากฏของคำย่อเบรลล์ กลุ่มตัวอักษรแต่ละกลุ่มจะใช้แทนคำแต่ละคำทำให้ไม่สามารถจะเขียนกลุ่มเบรลล์เหล่านี้ตามลำพังได้ ฉะนั้นเมื่อต้องการแสดงกลุ่มตัวอักษรเหล่านี้ในลักษณะที่ไม่ใช่ตัวย่อ ต้องใช้เครื่องหมายกำกับอักษร (⠠⠠) นำหน้ากลุ่มตัวอักษรเหล่านี้ ตัวอย่างเช่นเมื่อพบ ‘⠠⠠⠠⠠⠠’ (=ab) จะต้องตีความเป็น “about” แต่หากต้องการจะหมายถึงตัวอักษร “ab” จำเป็นต้องใส่เครื่องหมาย ‘⠠⠠’ ไว้ข้างหน้า **ซึ่งใช้บอกว่าสิ่งที่ตามมาจะเป็นตัวอักษร** เช่น ‘⠠⠠⠠⠠⠠’ จึงจะตีความเป็นตัวอักษร “ab” ได้



จากข้อมูลที่น่าเสนอมาข้างต้นพอจะสรุปได้ว่า ในการใช้ตัวย่อเบรลล์อังกฤษนั้นตัวย่อแต่ละตัวจะใช้แทนกลุ่มตัวอักษรที่แตกต่างกันจึงไม่ควรใช้ตัวย่อในตำแหน่งที่ไม่เหมาะสมเพราะจะทำให้เกิดความสับสน นอกจากนี้คำบางคำใช้ตัวย่อได้หลายรูปจึงควรเลือกกฎการใช้ตามลำดับความสำคัญเพื่อให้ได้รูปย่อที่เหมาะสมที่สุด สิ่งที่ต้องคำนึงถึงอีกประการคือต้องใช้ตัวย่อกับคำพยางค์เดียวกัน โดยเฉพาะในกรณีที่พยางค์นั้นเป็นส่วนหนึ่งของวิภัติ (prefix) หรือปัจจัย (suffix)

### 2.1.2 อักษรเบรลล์ไทย

มิสเจเนวีฟ คอลฟิลด์<sup>4</sup> ได้ร่วมกับนักวิชาการไทยกำหนดรหัสอักษรเบรลล์ไทยขึ้น (สมทรง พันธุ์สุวรรณ, 2538: 19; อาทิตยา บุญมาก และคณะ, 2547: 12) อย่างไรก็ตาม ในประเทศไทยยังไม่มีเอกสารหรืองานวิจัยชิ้นใดที่อธิบายอักษรเบรลล์ไทยอย่างเป็นระบบ มีเพียงการจัดการเรียนการสอนอักษรเบรลล์โดยการท่องจำรหัสอักษรเบรลล์เท่านั้น สำหรับอักษรเบรลล์ไทยได้มีผู้แสดงรหัสต่าง ๆ ทั้งพยัญชนะ สระ และวรรณยุกต์ ดังจะขอนำเสนอต่อไปนี้

#### 2.1.2.1 พยัญชนะในเบรลล์ไทย

อักษรเบรลล์ไทยประกอบด้วยพยัญชนะภาษาไทย 44 ตัว เช่นเดียวกับตัวอักษรปกติ ซึ่งสามารถแสดงรหัสในอักษรเบรลล์ได้ดังนี้

---

<sup>4</sup> มิสเจเนวีฟ คอลฟิลด์ (Genevive Caulfield) สุภาพสตรีตาบอดชาวอเมริกัน เป็นผู้ริเริ่มการศึกษาของคนตาบอดในประเทศไทย โดยร่วมกับนายแพทย์ ฝน แสงสิงแก้ว จัดตั้งโรงเรียนสอนคนตาบอดกรุงเทพ ขึ้นเมื่อปีพ.ศ.2482

ตารางที่ 2.19 พยัญชนะในเบรลล์ไทย

เบรลล์	แอสกี	อักษรไทย
⠠	g	ก
⠡	k	ข
⠢	Ok	ฅ
⠣	u	ค
⠤	-u	ก
⠥	,u	ฌ
⠦	]	ง
⠧	j	จ
⠨	/	ฉ
⠩	+	ช
⠪	!	ซ
⠬	,+	ฌ
⠭	,y	ญ
⠮	,d	ฎ
⠯	,\	ฏ
⠰	,t	ฐ
⠱	,)	ฑ
⠲	-)	ฒ
⠳	,n	ณ
⠴	d	ด
⠵	\	ต
⠶	t	ถ

เบรลล์	แอสกี	อักษรไทย
⠷	)	ท
⠸	0)	ฑ
⠹	n	น
⠺	v	บ
⠻	&	ป
⠼	p	ผ
⠽	x	ฝ
⠿	?	พ
⠾	\$	ฟ
⠿	,?	ภ
⠿	m	ม
⠿	y	ย
⠿	r	ร
⠿	l	ล
⠿	w	ว
⠿	,s	ศ
⠿	-s	ษ
⠿	s	ส
⠿	h	ห
⠿	,l	ฬ
⠿	o	อ
⠿	=	ฮ

สมทรง พันธุ์สุวรรณ (2538) ตั้งข้อสังเกตว่า พยัญชนะเบรลล์ไทยบางส่วน เกิดจากการนำเสียงพยัญชนะอังกฤษ 15 ตัวมาใช้เป็นฐานในการสร้างตัวอักษรเบรลล์ไทย ซึ่งประกอบด้วยตัวอักษรต่อไปนี้

ตารางที่ 2.20 ตัวอักษรเบรลล์ไทยพื้นฐาน 15 ตัว

อักษรไทย	เบรลล์ไทย	อักษรอังกฤษ	เบรลล์อังกฤษ
ก	⠠	g	⠠
ข	⠡	k	⠡
จ	⠢	j	⠢
ด	⠣	d	⠣
ต	⠤	t	⠤
น	⠥	n	⠥
พ	⠦	p	⠦
ม	⠧	m	⠧
ย	⠨	y	⠨
ร	⠩	r	⠩
ล	⠪	l	⠪
ว	⠬	w	⠬
ส	⠭	s	⠭
ห	⠮	h	⠮
อ	⠯	o	⠯

จากการทบทวนวรรณกรรม ผู้วิจัยยังไม่พบบงานเขียนใดๆ ที่อธิบายเกี่ยวกับเบรลล์ไทยในแง่มุมอื่น แต่ผู้วิจัยได้ค้นพบว่าในการสร้างพยัญชนะเบรลล์ไทย มีการนำพยัญชนะเบรลล์บางตัวมาเป็นฐาน เพื่อสร้างตัวพยัญชนะอื่นที่มีลักษณะใกล้เคียงกันใน 2 รูปแบบ ดังนี้

- รูปแบบที่ 1 การเติม ‘⠠’ (จุด 6) ในพยัญชนะฐานทำให้เกิดพยัญชนะใหม่ที่มีคุณสมบัติใกล้เคียงกันแต่เป็นอักษรที่ต่างหมู่กัน เช่น ‘⠠’ ใช้แทน “ส” ‘⠡’ ใช้แทน “ซ” พยัญชนะทั้งสองตัวนี้ต่างเป็นพยัญชนะเสียงเดียวกัน แต่ต่างกันตรงที่ “ส” เป็นอักษรสูง ในขณะที่ “ซ” เป็นอักษรต่ำ
- รูปแบบที่ 2 การเพิ่มเซลล์คือ ‘⠠’ (จุด 6) ‘⠡’ (จุด 3-6) หรือ ‘⠢’ (จุด 3-5-6) หน้าอักษรฐานทำให้เกิดพยัญชนะใหม่ที่มีเสียงใกล้เคียงกันและอยู่ในอักษรหมู่เดียวกัน เช่น ‘⠠’ ใช้แทน “ท” ‘⠡’ ใช้แทน “ท” ‘⠢’ ใช้แทน “ฒ” และ ‘⠣’ แทน “ธ” ซึ่งล้วนอยู่ในหมู่อักษรต่ำทั้งสิ้น

ข้อสังเกตเหล่านี้ทำให้ผู้วิจัยต้องการหาคำตอบว่า พยัญชนะเบรลล์ไทยมีรูปที่ปรากฏอย่างเป็นระบบหรือไม่ และพยัญชนะแต่ละตัวมีความสัมพันธ์กันในลักษณะใด วิทยานิพนธ์ฉบับนี้จึงมุ่งที่จะหาคำตอบให้กับประเด็นเหล่านี้

#### 2.1.2.2 สระและวรรณยุกต์ในเบรลล์ไทย

นอกจากพยัญชนะแล้ว อักษรเบรลล์ไทยยังประกอบด้วยสระและวรรณยุกต์ดังนี้

ตารางที่ 2.21 สระและวรรณยุกต์ในเบรลล์ไทย

เบรลล์	แอสกี	อักษรปกติ
⠁	a	ะ
⠃	>	ั
⠅	*	า
⠇	z	ำ
⠉	b	ิ
⠋	2	ี
⠍	{	ึ
⠏	5	ื
⠑	c	ุ
⠓	3	ู
⠗	r1	ฤ
⠗⠗	r1*	ฤา
⠗⠗	l1	ฦ
⠗⠗⠗	l1*	ฦา
⠙	f	เ
⠛	<	แ
⠝	i	โ
⠟	:	ใ

เบรลล์	แอสกี	อักษรปกติ
⠠	1	า
⠠⠠	:1	ใ
⠠	,	ี
⠠	9	อ
⠠	4	็
⠠	7	ื
⠠	8	ั
⠠	0	็
⠠⠠	fa	เะ
⠠⠠	<a	แะ
⠠⠠	ia	โะ
⠠⠠	oa	เาะ
⠠	o	อ
⠠⠠	%a	เอะ
⠠	%	เื่อ
⠠⠠	(a	เียะ
⠠	(	เีย
⠠⠠	qa	เื่อะ
⠠	q	เื่อ
⠠⠠	ea	ัวะ
⠠	e	ัว
⠠	6	เ่า

จากรหัสอักษรเบรลล์ไทยที่นำเสนอมาข้างต้น ผู้วิจัยพบว่ามีการนำอักษรเบรลล์รูปหนึ่งมาดัดแปลงเพื่อทำให้เป็นอักษรอีกรูปหนึ่งที่มีลักษณะใกล้เคียงกัน การดัดแปลงทำได้ 2 ลักษณะคือ ลักษณะแรก นำสระเบรลล์รูปหนึ่งมาเติม ‘⠠’ (จุด 6) เพื่อทำให้เป็นสระอีกรูปหนึ่งที่มีลักษณะ



เบรลล์ เพื่อใช้เป็นฐานความรู้ในการสร้างระบบการถ่ายถอดอักษรที่รองรับกับการแปลงตัวเลขในข้อความเบรลล์สองภาษาที่ปนกัน

ในการเขียนตัวเลขเบรลล์ ทำได้โดยใช้ตัวอักษร 10 ตัวแรกในภาษาอังกฤษมาใช้แทนตัวเลขตั้งแต่ 1 - 0 ตามลำดับ ดังแสดงในตารางต่อไปนี้ (Braille Authority of North America, 2002: VII-1)

ตารางที่ 2.22 ตัวเลขในอักษรเบรลล์

เบรลล์	⠠	⠡	⠢	⠣	⠤	⠥	⠦	⠧	⠨	⠩
แอสกี	a	b	c	d	e	f	g	h	i	j
ตัวเลข	1	2	3	4	5	6	7	8	9	0

#### 2.1.3.1 วิธีการเขียนตัวเลขในอักษรเบรลล์

สัญลักษณ์ตัวเลขที่ปรากฏในตารางข้างต้น มีวิธีการใช้ดังนี้ (Risjord, 2009: 29-33; California Community Colleges, 2010: 32-36)

(1) การเขียนตัวเลขโดด ให้ใช้เครื่องหมาย ‘⠠’ (= #) หรือเครื่องหมายนำเลข นำหน้าตัวอักษรเหล่านั้น เพื่อแสดงว่าเป็นตัวเลข เช่น

⠠⠠ (= #a) ใช้แทน ‘1’

⠠⠡ (= #b) ใช้แทน ‘2’

⠠⠢ (= #c) ใช้แทน ‘3’

(2) หากเลขจำนวนใดมีตัวเลขมากกว่า 1 ตัว ให้ใช้เครื่องหมายนำเลขนำหน้าตัวเลขตัวหน้า

⠠⠠⠠ (= #aj) ใช้แทน ‘10’

⠠⠠⠡ (= #ajj) ใช้แทน ‘100’

⠠⠠⠢ (= #ajjj) ใช้แทน ‘1000’

(3) การเขียนตัวเลขกับคำที่ปรากฏข้างหลัง

ด้วยเหตุที่เบรลล์อังกฤษ 10 ตัวแรกได้นำมาใช้แทนตัวเลขเบรลล์ จึงทำให้จำเป็นต้องเขียนเว้นวรรคระหว่างตัวเลขและตัวอักษรเสมอ เช่น ‘⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠’ (= #aj bags) หรือ “10 bags” ในอักษรปกติ หากเขียนชิดกันอาจทำให้เกิดการตีความผิดเป็น “10217s” ได้ เนื่องจาก ‘⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠⠠’ (= #ajjj) อยู่ชิดกับตัวเลขทำให้เกิดความสับสนได้ ดังนั้นเครื่องหมายวรรคจึงเป็นตัวบ่งชี้จุดสิ้นสุดของตัวเลขจำนวนใดจำนวนหนึ่ง

## (4) การเขียนตัวเลขและตัวอักษรปนกัน

นอกจากการใช้เครื่องหมายเว้นวรรคแล้ว ยังมีการใช้ ‘::’ (=) หลังตัวเลข เพื่อแยกตัวเลขและตัวอักษรออกจากกันในกรณีที่ต้องเขียนตัวอักษรหลังตัวเลขโดยไม่มีเว้นวรรค เช่น ‘::: :::’ (= #b :: a) จะใช้แทน “2a” เป็นต้น ในทางตรงกันข้ามหากต้องการเขียนตัวเลขหลังตัวอักษร ก็ต้องใช้เครื่องหมาย ‘::’ (= #) เช่น ‘::: :::’ (= b :: a) ใช้แทน “b1”

## (5) การใช้เครื่องหมายที่ปรากฏกับตัวเลข

หากมีเครื่องหมายเหล่านี้ปรากฏในสายอักขระของตัวเลขให้ใช้ ‘::’ (= #) ที่ตำแหน่งต้นของตัวเลขเพียงครั้งเดียว เครื่องหมายเหล่านี้แสดงได้ดังนี้

ตารางที่ 2.23 เครื่องหมายที่ปรากฏกับตัวเลข

สัญลักษณ์เบอร์ลด์	สัญลักษณ์แอสกี	ใช้แทนเครื่องหมาย	หมายเหตุ
::	+	+	(บวก)
::	-	-	(ลบ)
:::	@*	*	(คูณ)
::	/	÷	(หาร)
:::	./	/	(หาร หรือทับ)
::	.	.	(จุดทศนิยม)
::	.	.	(จุดท้ายตัวเลข)
::	1	,	(คอมม่า)
::	3	:	(โคลอน)
:::	@3p	%	(เปอร์เซ็นต์)

ตัวอย่างเช่น หากต้องการเขียน 2-3 (สองถึงสาม) ในอักษรเบอร์ลด์สามารถทำได้ดังนี้

ถูก: ‘::: :::’ (#b-c)

ผิด: ‘::: ::: :::’ (#b-#c)

จะเห็นว่าระบบการเขียนตัวเลขเบอร์ลด์ที่ได้นำเสนอนี้เป็นเพียงสัญลักษณ์ที่ปรากฏในข้อความทั่วไปในอักษรเบอร์ลด์ อย่างไรก็ตามในศาสตร์ต่าง ๆ ที่เกี่ยวข้องกับตัวเลข เช่น คณิตศาสตร์ วิทยาศาสตร์ คนตรี ฯลฯ จะมีวิธีการเขียนที่แตกต่างออกไป เช่น การใช้เลขต่ำเพื่อแทนตัวเลขเหล่านี้



เนื่องจากวิทยานิพนธ์นี้มุ่งที่จะพัฒนาระบบการแปลงอักษรเบรลล์ที่ปรากฏในข้อความทั่วไป ดังนั้นงานวิจัยฉบับนี้จะไม่ขอก้าวถึงระบบตัวเลขที่ใช้เฉพาะด้านที่ได้อภิปรายมาข้างต้น

## 2.2 การจำแนกภาษาด้วยคอมพิวเตอร์

สิ่งที่ได้กล่าวมาแล้วว่าวัตถุประสงค์สำคัญสำหรับวิทยานิพนธ์ฉบับนี้คือ การถ่ายถอดอักษรเบรลล์ที่มีภาษาไทยและอังกฤษปนกันเป็นอักษรปกติ อย่างไรก็ตามเป็นไปได้ที่คอมพิวเตอร์จะให้ผลลัพธ์ได้ถูกต้องหากไม่รู้ว่าข้อความเบรลล์ที่ต้องการถ่ายถอดเป็นภาษาใด เนื่องจากเบรลล์อังกฤษและเบรลล์ไทยต่างก็ใช้ตัวอักษรระบบ 6 จุด จึงเกิดการซ้ำรูปในภาษาทั้งสอง ทำให้ยากที่เครื่องคอมพิวเตอร์จะบอกได้ว่ากำลังทำงานกับอักษรเบรลล์ของภาษาใด

ดังนั้นก่อนที่จะทำการถ่ายถอดอักษรเบรลล์เป็นอักษรปกติ คอมพิวเตอร์ต้องจำแนกภาษาในข้อความเบรลล์ให้ได้เป็นอันดับแรก

ในการประมวลผลภาษาธรรมชาติ การจำแนกภาษาด้วยคอมพิวเตอร์ จัดเป็นประเด็นปัญหาหนึ่งในการจัดประเภทข้อความ (text classification) เป็นการระบุว่าข้อความที่คอมพิวเตอร์กำลังประมวลผลอยู่นั้นเป็นภาษาใด (Grothe, De Luca & Nürnberger 2008: 1) การจำแนกภาษาด้วยคอมพิวเตอร์ประกอบด้วย 2 ส่วนหลัก คือ แบบจำลองการจำแนกภาษา (language identification models) และวิธีการในการจำแนกภาษา (language classification method)

ส่วนสำคัญในแบบจำลองการจำแนกภาษาคือ แบบจำลองภาษา (language models) ซึ่งอาจจะเป็นเซตของคำหรือเซตของกลุ่มตัวอักษรซึ่งมีลักษณะพิเศษที่ได้มาจากคลังข้อมูล โดยนำมาใช้เป็นต้นแบบในการเปรียบเทียบกับข้อความที่ต้องการระบุภาษา หากพิจารณาตามลักษณะข้อมูลที่ปรากฏในแบบจำลองภาษาสามารถจะแบ่งวิธีการจำแนกภาษาออกได้เป็น 2 ประเภท คือ การจำแนกภาษาในระดับคำ (word based) และการจำแนกภาษาในระดับอักขระ (character based)

ต่อไปนี้ ผู้วิจัยจะกล่าวถึงการจำแนกภาษาในระดับคำ และการจำแนกภาษาในระดับอักขระ โดยมีรายละเอียดดังนี้

### 2.2.1 การจำแนกภาษาในระดับคำ

ในการจำแนกภาษาหากใช้คำมาสร้างแบบจำลองภาษา จากการทบทวนวรรณกรรมด้านนี้พบว่าวิธีการทำได้แตกต่างกันอย่างน้อย 2 วิธี วิธีแรกคือการใช้คำที่มีความถี่ในการปรากฏสูงในคลังข้อมูล (frequent word-based approach) มาเป็นตัวแทนภาษานั้นๆ วิธีที่สองคือการใช้คำขนาด

สั้น ซึ่งจะมีจำนวนอักขระไม่เกินจำนวนที่กำหนด (short word-based approach) มาเป็นตัวแทนภาษานั้นๆ

### 2.2.1.1 การใช้คำที่มีความถี่สูง

การจำแนกภาษาด้วยวิธีการใช้คำที่มีความถี่สูง อิงกับแนวคิดที่ว่า ในแต่ละภาษามีคำจำนวนหนึ่งที่มีความถี่ในการปรากฏสูงและมีรูปแตกต่างจากภาษาอื่น คำเหล่านี้สามารถนำมาใช้ในการสร้างแบบจำลองภาษา (language models) เพื่อบ่งชี้แต่ละภาษาได้

ในการจำแนกภาษาด้วยวิธีการนี้ เริ่มต้นจากการแจกแจงรายการคำที่มีความถี่ในการปรากฏสูงที่สุดจำนวนหนึ่งในคลังข้อมูลของแต่ละภาษา แล้วนำคำเหล่านั้นมาสร้างแบบจำลองของแต่ละภาษา จากนั้นจึงนำคำแต่ละคำในข้อความที่ต้องการระบุภาษาไปเปรียบเทียบกับค่าความคาดหมายว่าคำเหล่านั้นใกล้เคียงกับแบบจำลองภาษาใดมากที่สุด หลังจากทำเช่นนี้จนครบทั้งข้อความ ก็จะระบุให้ข้อความนำเข้าไปเป็นภาษาที่คำนวณแล้วได้คะแนนรวมสูงสุด

ซูเตอร์ (Souter et al., 1994) มาร์ติโน (Martino et al., 2001) และโควีย์ (Cowie et al., 1999) ต่างก็นำแนวคิดนี้มาใช้ในการจำแนกภาษา โดยที่ซูเตอร์และมาร์ติโนต้องการจำแนกข้อความของภาษาที่ใช้ในทวีปยุโรป ในขณะที่โควีย์ต้องการจำแนกภาษาที่ใช้ในทวีปยุโรป เอเชีย และแอฟริกา สำหรับการสร้างแบบจำลองภาษา ซูเตอร์และมาร์ติโนได้ใช้คำที่มีความถี่สูงสุด 100 คำแรกของแต่ละภาษา ในขณะที่โควีย์ใช้คำที่มีความถี่สูงถึง 1,000 คำแรก

สิ่งที่น่าสังเกตคือผู้วิจัยทั้ง 3 คน มีวิธีการคิดคะแนนของแบบจำลองที่ต่างกัน ซูเตอร์และโควีย์ใช้วิธีการเพิ่มค่าตัวนับ (counter) ทีละ 1 ให้กับแบบจำลองภาษาที่มีค่าตรงกับคำในข้อความนำเข้าไป ในขณะที่มาร์ติโนใช้วิธีการเพิ่มค่าตัวนับ (ซึ่งมาร์ติโนเรียกว่า accumulator) ให้กับแต่ละแบบจำลองภาษา โดยใช้ค่าความถี่ปรับแบบบรรทัดฐาน (normalized frequency of occurrence หรือ NFO) ซึ่งคำนวณโดย นำความถี่ของคำแต่ละคำในแต่ละแบบจำลองมาหารด้วยจำนวนคำทั้งหมดในคลังข้อมูลภาษานั้น แล้วนำผลลัพธ์ทุกตัวมารวมกันหารด้วยความถี่ของคำที่มีการปรากฏสูงสุดในแบบจำลองนั้น

นอกจากการนำคำที่มีความถี่สูงสุดมาใช้จำแนกภาษาของข้อความได้แล้ว ค่าขนาดสั้นที่ปรากฏในแต่ละภาษาก็ยังสามารถใช้เป็นตัวบ่งชี้ภาษาได้ ดังจะได้นำเสนอในหัวข้อต่อไป

### 2.2.1.2 การใช้คำขนาดสั้น

การใช้คำขนาดสั้นเพื่อจำแนกภาษาเป็นการรวบรวมคำที่มีจำนวนอักขระไม่เกินที่กำหนดไว้มาสร้างเป็นแบบจำลองภาษาของแต่ละภาษาโดยไม่คำนึงถึงความถี่ในการปรากฏของคำเหล่านั้น

การจำแนกภาษาด้วยวิธีนี้ใช้แนวคิดที่ว่า คำทั่วไป (common word) ที่ทำให้สามารถระบุภาษาได้ ได้แก่ คำหน้าที่ (function) เช่น คำสันธาน คำบุพบท ส่วนใหญ่เป็นคำที่มีความยาวไม่มากนัก (Grefenstette, G. 1995:3)

ในการจำแนกภาษาด้วยวิธีนี้ เริ่มจากนำข้อความของแต่ละภาษาจากคลังข้อมูล ไปแบ่งคำ แล้วเลือกเฉพาะคำที่มีจำนวนอักขระไม่เกินจำนวนที่กำหนดไว้ไปสร้างแบบจำลองภาษา หลังจากนั้นนำคำเหล่านี้ไปคำนวณความน่าจะเป็น (probability) โดยคิดจาก ความถี่ในการเกิดของแต่ละคำหารด้วยผลรวมความถี่ของคำทั้งหมดในแบบจำลองภาษา ในการจำแนกภาษาของประโยคที่นำเข้ามา โดยนำคำแต่ละคำในประโยคไปเปรียบเทียบกับคำในแต่ละแบบจำลองภาษา หากตรงกับคำในแบบจำลองใดก็ให้นำค่าความน่าจะเป็นของคำนั้น (probability) ในแบบจำลองไปรวมกับผลรวมเดิมของแบบจำลองภาษานั้น หากไม่ตรงกับคำในแบบจำลองใดก็ให้นำค่าความน่าจะเป็นต่ำสุด (minimum probability) ที่กำหนดขึ้น ไปรวมกับผลรวมเดิมของแบบจำลองภาษานั้นเช่นกัน หลังจากเปรียบเทียบคำทุกคำกับแบบจำลองภาษาทั้งหมด แล้วจึงทำการเปรียบเทียบผลรวมค่าความน่าจะเป็นของประโยคที่ได้ในแต่ละแบบจำลองภาษา เพื่อหาแบบจำลองภาษาที่มีคะแนนรวมสูงสุด อันจะทำให้ระบุได้ว่าเป็นภาษาใด

เกรเฟนส์เตท (Grefenstette, G. 1995) และเพรเกอร์ (Prager, G.M. 1999) ได้ใช้วิธีการดังกล่าวมาจำแนกภาษา โดยเกรเฟนส์เตทเลือกเฉพาะคำที่มีความยาวไม่เกิน 5 ตัวอักษรซึ่งส่วนใหญ่เป็นคำหน้าที่ (function words) มาสร้างแบบจำลองภาษา ในขณะที่เพรเกอร์ใช้ความยาวของคำไม่เกิน 4 ตัวอักษร เนื่องจากเชื่อว่าคำที่มีความยาวเท่าที่ระบุมา ก็สามารถจำแนกภาษาได้ดีเท่ากับการใช้คำที่มีขนาด ตามที่เกรเฟนส์เตทได้นำเสนอ

เป็นที่น่าสังเกตว่า การจำแนกภาษาในระดับคำมีข้อจำกัดบางประการ ประการแรก ในบางกรณีข้อความที่นำมา ไม่มีคำสันหรือคำที่มีความถี่สูงปรากฏอยู่เลย เช่น ข้อความพาดหัวข่าว ซึ่งหากพบกรณีนี้คอมพิวเตอร์จะไม่สามารถระบุภาษาของข้อความเหล่านั้นได้ ประการต่อมา ภาษาต่าง ๆ ในโลก เช่น ภาษาไทย จีน เกาหลี ญี่ปุ่น ฯลฯ ไม่มีการเว้นวรรคระหว่างคำ หากจะนำวิธีการดังกล่าวมาใช้กับภาษาเหล่านี้ ก็จำเป็นต้องทำการตัดคำลวงหน้า ทำให้เกิดความซับซ้อนยิ่งขึ้น จึงไม่เหมาะสมที่จะนำวิธีการจำแนกภาษาในระดับคำมาใช้กับภาษาที่มีข้อจำกัดดังกล่าว

## 2.2.2 การจำแนกภาษาในระดับอักขระ

ภาษาต่าง ๆ ในโลกอาจใช้ตัวอักษรร่วมกัน เช่นภาษาอังกฤษ เยอรมัน ฝรั่งเศส แต่ภาษาเหล่านี้ก็มีการจัดเรียงตัวอักษรที่แตกต่างกัน เมื่อพิจารณารูปแบบการจัดเรียงตัวอักษรก็อาจจะระบุได้ว่าเป็นภาษาใด เช่นคำที่ลงท้าย -ck สามารถระบุได้ว่าเป็นภาษาอังกฤษ ในขณะที่คำที่ลงท้าย -ez จะ

สามารถระบุได้ว่าเป็นภาษาฝรั่งเศส (Grefenstette, G. 1995: 2) ในการจำแนกภาษาในระดับอักขระ อาจทำได้ด้วยวิธีการต่าง ๆ คือ การใช้สายอักขระที่มีความเป็นเอกลักษณ์ของแต่ละภาษา (unique character strings) และการใช้แบบจำลองเอ็นแกรม (n-grams model) มาใช้ระบุภาษา

#### 2.2.2.1 การใช้สายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา

การจำแนกภาษาโดยใช้สายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา ตั้งอยู่บนสมมติฐานที่ว่า ภาษาต่าง ๆ ในโลกจะมีการเรียงตัวของสายอักขระที่ไม่เหมือนกันแม้จะเป็นภาษาในตระกูลเดียวกันก็ตาม ความแตกต่างในการเรียงตัวของสายอักขระ เป็นลักษณะพื้นฐานของแต่ละภาษาที่สามารถนำมาใช้ในการบ่งชี้เอกลักษณ์ของภาษานั้น ๆ ได้ เช่นคำที่ขึ้นต้นด้วย LL จะเป็นภาษาเวลส์ (Welsh), หรือคำที่มี CZY ปรากฏนั้นจะเป็นภาษาโปแลนด์ (Polish) เป็นต้น

การหาสายอักขระที่เป็นเอกลักษณ์เฉพาะแต่ละภาษาอาจทำได้โดยอาศัยนักภาษาศาสตร์ที่มีความเชี่ยวชาญในภาษาเหล่านั้น หรือนำข้อความที่สุ่มจากภาษาต่าง ๆ มาเปรียบเทียบกับเพื่อหาสายอักขระที่ไม่ซ้ำกัน (Dunning, T. 1994: 3) อย่างไรก็ตาม ในการหาสายอักขระที่ปรากฏเฉพาะแต่ละภาษานั้น อาจจะต้องกำหนดความยาวของสายอักขระที่เหมาะสมจึงจะสามารถจำแนกภาษาได้ถูกต้อง (Souter et al., 1994: 187) ในบางกรณี สายอักขระที่มีความยาว 2 ตัว อาจไม่สามารถนำมาใช้จำแนกภาษาหนึ่งออกจากอีกภาษาหนึ่งได้ เนื่องจากภาษาทั้งสองนั้นมีการเรียงตัวของอักขระ 2 ตัวที่ซ้ำกันเป็นจำนวนมากทำให้จำเป็นต้องใช้อักขระที่มีความยาวมากกว่านั้นจึงจะสามารถระบุภาษาได้

ในการจำแนกภาษาของข้อความนำเข้าด้วยวิธีนี้ ทำได้โดยนำสายอักขระของข้อความนำเข้าไปเปรียบเทียบกับสายอักขระในแต่ละแบบจำลองภาษา แล้วให้ค่าคะแนนสำหรับแต่ละแบบจำลองภาษา แบบจำลองใดได้ค่าคะแนนมากที่สุดก็จะระบุให้ข้อความนำเข้าเป็นภาษานั้น

ชูเตอร์ (1994) ได้ตั้งข้อสังเกตว่า วิธีดังกล่าวให้ความถูกต้องในการจำแนกภาษาได้น้อยกว่าการใช้แบบจำลองเอ็นแกรมและการใช้คำที่มีความถี่สูง เนื่องจากบางครั้งคลังข้อความที่นำมาใช้ในการสร้างแบบจำลองภาษามีข้อมูลที่ไม่ครอบคลุมกับภาษานั้น ๆ เช่นในคลังข้อมูลที่ชูเตอร์นำมาใช้ มีการปรากฏของ  $nx$  ในคลังข้อมูลภาษาโปรตุเกสแต่ไม่ปรากฏในคลังข้อมูลภาษาอังกฤษ ทำให้ไม่สามารถระบุได้ว่า “anxious” เป็นคำในภาษาอังกฤษได้

### 2.2.2.2 การใช้แบบจำลองเอ็นแกรม

ในการจำแนกภาษาในระดับอักขระอาจทำได้โดยการใช้แบบจำลองเอ็นแกรม แบบจำลองเอ็นแกรมคือแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของสายอักขระ (character sequence) ที่เรียงต่อกันจำนวน  $n$  ตัว โดยค่าความน่าจะเป็นของชุดอักขระสามารถคำนวณได้จากคลังข้อมูลที่ได้รวบรวมไว้สำหรับแต่ละภาษา (Grothe, De Luca & Nürnbergger, 2008: 2)

สิ่งที่จำเป็นสำหรับการจำแนกภาษาโดยใช้เทคนิคเอ็นแกรม คือแบบจำลองภาษา (language models) ของแต่ละภาษา และแบบจำลองข้อความ (document models) โดยที่แต่ละแบบจำลองภาษา เกิดจากการรวบรวมสายอักขระเอ็นแกรมจากคลังข้อความของแต่ละภาษา ในขณะที่แบบจำลองข้อความ เกิดจากการรวบรวมสายอักขระเอ็นแกรมจากข้อความนำเข้า ในการระบุภาษาของข้อความใดข้อความหนึ่ง ทำได้โดยนำแต่ละสายอักขระในแบบจำลองข้อความไปเปรียบเทียบกับสายอักขระในแต่ละแบบจำลองภาษา เพื่อหาความน่าจะเป็นในการปรากฏของสายอักขระนั้น ในแต่ละแบบจำลองภาษา หลังจากนั้นค่าความน่าจะเป็นของสายอักขระทั้งหมดสำหรับแต่ละแบบจำลองภาษาจะถูกนำไปเปรียบเทียบกันเพื่อบ่งชี้ว่าข้อความนำเข้าจัดอยู่ในภาษาใด

เกรเฟนส์เตท(Grefenstette, G. 1995) ได้ใช้โปรแกรมในการจำแนกภาษาของข้อความโดยสร้างแบบจำลองภาษาที่ประกอบด้วยไตรแกรมที่มีความถี่ในการปรากฏในลำดับแรก ๆ ของคลังข้อความ ECI (European Corpus Initiative) ในการจำแนกภาษา ทำโดยนำแต่ละไตรแกรมของแบบจำลองข้อความ ซึ่งได้มาจากประโยคที่นำเข้าแต่ละประโยคมาเปรียบเทียบกับไตรแกรมในแต่ละแบบจำลองภาษา หากพบไตรแกรมที่เหมือนกันจะนำค่าความน่าจะเป็น (probability) ของไตรแกรมนั้น (ซึ่งคำนวณจากความถี่ของไตรแกรมนั้นหารด้วยความถี่ของไตรแกรมทุกตัวในแบบจำลองภาษา) มาเพิ่มคะแนนให้กับแบบจำลองภาษาที่พบไตรแกรม แต่หากไม่พบก็จะเพิ่มค่าความน่าจะเป็นที่น้อยที่สุด (minimal probability) ที่กำหนดขึ้นให้กับแบบจำลองภาษาที่ไม่พบไตรแกรมเช่นกัน หลังจากเปรียบเทียบทุกไตรแกรมในแบบจำลองข้อความกับไตรแกรมในแบบจำลองภาษาทั้งหมดแล้ว จะนำผลรวมค่าความน่าจะเป็นของแต่ละแบบจำลองภาษามาเปรียบเทียบกันเพื่อหาความน่าจะเป็นสูงสุด อันจะทำให้ระบุได้ว่าประโยคที่นำเข้านั้นเป็นภาษาใด

คาฟนาร์และเทรนเคิล (Cavnar, W. B. and Trenkle, J. M. 1994) ได้ใช้วิธีการจำแนกภาษาโดยใช้เอ็นแกรมของสายอักขระเช่นกัน เพียงแต่วิธีการนับคะแนนที่ต่างไป โดยไม่ได้นับจากค่าความน่าจะเป็นของสายอักขระแต่ใช้การจัดลำดับ (ranking) ของสายอักขระเอ็นแกรมของแต่ละแบบจำลองภาษา วิธีการระบุภาษาคือการนำแต่ละสายอักขระในแบบจำลองข้อความไปเปรียบเทียบกับสายอักขระดังกล่าวในแต่ละแบบจำลองภาษาหรือไม่ หากพบสายอักขระที่เหมือนกันในแบบจำลองภาษาใดจะนำลำดับของสายอักขระในแบบจำลองข้อความและแบบจำลองภาษานั้นมา

เปรียบเทียบกันเพื่อหาระยะห่างลำดับ (เช่นสายอักขระxxx ปรากฏในลำดับที่ 6 ในแบบจำลองภาษาอังกฤษแต่ปรากฏในลำดับที่ 2 ในแบบจำลองข้อความ ก็จะมีระยะห่างลำดับเป็น  $6-2=4$ ) แล้วนำค่าระยะห่างลำดับที่ได้ไปรวมกับค่าระยะห่างลำดับที่มีอยู่เดิมของแบบจำลองภาษานั้น หากไม่พบสายอักขระที่เหมือนกันในภาษาใดจะกำหนดค่าระยะห่างลำดับให้เป็นตัวเลขกลางซึ่งมีค่าสูงสุด (default maximum value) แล้วนำค่านี้ไปรวมกับค่าระยะห่างลำดับที่มีอยู่เดิมของแบบจำลองภาษานั้นเช่นกัน แบบจำลองภาษาใดที่มีผลรวมค่าระยะห่างลำดับต่ำที่สุดก็จะระบุให้ข้อความนำเข้าเป็นภาษานั้น

จากสิ่งที่ได้อภิปรายข้างต้น ผู้วิจัยพบว่าวิธีที่หลากหลายในการจำแนกภาษาด้วยคอมพิวเตอร์ บางวิธีก็เหมาะที่จะประยุกต์ใช้กับงานวิจัยที่ผู้วิจัยกำลังทำการศึกษา ดังนั้นผู้วิจัยจะขออภิปรายถึงวิธีการที่เหมาะสมที่จะนำมาใช้ในการระบุภาษาในอักษรเบรลล์ ดังนี้

ดังที่ได้กล่าวมาแล้วว่า วิธีการหนึ่งที่ยอมรับนำมาใช้ระบุภาษาของข้อความ คือการใช้คำจำนวนหนึ่งที่แตกต่างกันในแต่ละภาษามาจำแนกข้อความของภาษาเหล่านั้นออกจากกัน (ดูหัวข้อ 2.2.2.1) แต่ผู้วิจัยพบว่า ในการจำแนกเบรลล์ไทยออกจากเบรลล์อังกฤษนั้น ไม่สามารถจะใช้คำแต่ละคำมาจำแนกข้อความเบรลล์ของสองภาษานี้ออกจากกันได้ เนื่องจากข้อความภาษาไทยไม่มีการระบุขอบเขตของคำอย่างชัดเจน (กานดา รุณนะพงศา และปโยธร อุราธรรมกุล. 2549: 3) แต่ด้วยเหตุนี้ ข้อความแต่ละวรรคในภาษาอังกฤษประกอบด้วยคำเพียงคำเดียว ในขณะที่ภาษาไทยเกิดจากการนำคำหลายคำมาเขียนต่อกัน จึงทำให้ข้อความภาษาอังกฤษมีขนาดสั้นกว่าข้อความภาษาไทยอย่างชัดเจน ดังนั้นจึงสามารถจำแนกภาษาไทยออกจากภาษาอังกฤษได้โดยการพิจารณาความยาวสายอักขระ โดยที่สายอักขระใดมีความยาวมากก็น่าจะระบุได้ว่าเป็นภาษาไทย

นอกจากนี้ ยังสามารถนำวิธีการจำแนกภาษาโดยใช้สายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษามาใช้ระบุภาษาในอักษรเบรลล์ได้ แม้ว่าเบรลล์ไทยและเบรลล์อังกฤษจะมีรูปตัวอักษรที่เหมือนกัน ทั้งสองภาษานี้ก็มีรูปแบบของการเรียงตัวของตัวอักษรที่แตกต่างกันเป็นจำนวนมาก เช่น ‘::: ::’ เป็นภาษาไทย (=กิน) ไม่ใช่ภาษาอังกฤษ(=gbn) ‘:: :::: ::’ เป็นภาษาอังกฤษ (= house) ไม่ใช่ภาษาไทย (= หอคั่ว) อย่างไรก็ตาม ในบางครั้งอาจมีการเรียงต่อกันของตัวอักษรที่เหมือนกันในสองภาษา เช่น ‘:: ::::’ ซึ่งอาจใช้แทนคำว่า “dog” หรือ “ดอก” ในภาษาไทยก็ได้ หากเกิดกรณีนี้ขึ้นจะส่งผลให้ไม่สามารถระบุภาษาได้ จึงจำเป็นต้องนำสถิติความน่าจะเป็นมาใช้ร่วมกับการระบุภาษาด้วยสายอักขระที่เป็นเอกลักษณ์ของแต่ละภาษา

สำหรับรายละเอียดเกี่ยวกับแบบจำลองการระบุภาษาในอักษรเบรลล์ ที่จะนำเสนอในวิทยานิพนธ์ฉบับนี้ ผู้วิจัยจะอภิปรายในหัวข้อที่เกี่ยวข้องต่อไป

## 2.3 การถ่ายถอดอักษร

ดังที่ได้นำเสนอมาแล้วว่า วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ที่จะถ่ายถอดอักษรเบรลล์ที่มีการปนกันระหว่างภาษาไทยและภาษาอังกฤษให้เป็นอักษรปกติด้วยคอมพิวเตอร์ ซึ่งกระบวนการนี้มีลักษณะคล้ายคลึงกับการถ่ายถอดข้อความจากระบบตัวเขียนหนึ่งไปสู่อีกระบบตัวเขียนหนึ่ง ด้วยเหตุนี้ผู้วิจัยจึงได้ทำการทบทวนวรรณกรรมที่เกี่ยวข้องเพื่อจะนำมาประยุกต์ใช้ในการถ่ายถอดอักษรเบรลล์เป็นอักษรปกติ โดยหัวข้อที่ 2.4.1 จะนำเสนอ แนวคิดทั่วไปเกี่ยวกับการถ่ายถอดอักษร หลังจากนั้นหัวข้อที่ 2.4.2 จะนำเสนอแนวทางการถ่ายถอดอักษรด้วยคอมพิวเตอร์

### 2.3.1 แนวคิดทั่วไปเกี่ยวกับการถ่ายถอดอักษร

การถ่ายถอดอักษร (transliteration) เป็นการบันทึกตัวเขียนในภาษาหนึ่ง (ภาษาต้นทาง) ด้วยตัวเขียนที่สอดคล้องกันในอีกภาษาหนึ่ง (ภาษาปลายทาง) โดยที่เสียงของทั้งสองภาษายังคงมีความใกล้เคียงกัน (Knight, K. & Greal, J. 1997: 599 ; Catford, J. Cunnison 1965: 50) ตัวอย่างเช่น คำว่า “ถ่ายเสียง” สามารถถ่ายถอดอักษรในภาษาอังกฤษได้เป็น “taisiang” เป็นต้น

หากจำแนกการถ่ายถอดอักษรโดยยึดความสอดคล้องของภาษาต้นทางและภาษาปลายทาง ก็อาจแบ่งการถ่ายถอดอักษรออกได้เป็น 2 ประเภทคือ การถ่ายถอดอักษรโดยยึดอักษรเป็นสำคัญ และ การถ่ายถอดอักษรโดยยึดเสียงเป็นสำคัญ (Coulmas, Folrian 1992: 241-243)

ในประเภทแรก การถ่ายถอดอักษรโดยยึดอักษรเป็นสำคัญเป็นการถอดรูปอักษรที่ปรากฏในคำนั้น ๆ มาให้ครบทุกรูปในลักษณะอักษร (หรือกลุ่มตัวอักษร) ต่ออักษร (หรือกลุ่มตัวอักษร) โดยไม่คำนึงว่าจะออกเสียงผิดเพี้ยนไปจากคำเดิมหรือไม่ การถ่ายถอดอักษรในลักษณะนี้ใช้ในกรณีที่ต้องการคงรูปศัพท์ในคำดั้งเดิมเอาไว้ เช่น การถ่ายถอดชื่อบุคคล (เช่น บงกช = Bonggoch) และการถ่ายถอดชื่อทางภูมิศาสตร์ต่าง ๆ (เช่น สุวรรณภูมิ = Suvarnabhumi) เป็นต้น จะเห็นได้ว่า การถ่ายถอดอักษรลักษณะนี้มีข้อดีคือ สามารถถอดอักษรเพื่อใช้ในงานค้นคืนข้ามภาษา (cross language information retrieval) ได้ง่าย เนื่องจากเป็นการจับคู่ตัวอักษรหรือกลุ่มตัวอักษรในภาษาหนึ่งกับตัวอักษรหรือกลุ่มตัวอักษรอีกภาษาหนึ่งโดยตรง แต่ก็มีข้อเสียคือ คำที่ได้จากการถ่ายถอดอักษรในภาษาปลายทางจะออกเสียงผิดเพี้ยนไปจากคำในภาษาต้นทาง เช่น ชื่อ Bonggoch ถ้ายึดตามการออกเสียงแบบภาษาอังกฤษจะต้องออกเสียงว่า “บงกช” ซึ่งมีเสียงแตกต่างจากคำว่า “บงกช” ที่เป็นคำดั้งเดิมในภาษาไทย

ในอีกประเภทหนึ่ง การถ่ายถอดอักษรโดยยึดเสียงเป็นสำคัญ เป็นการถ่ายถอดที่ไม่จำเป็นต้องถอดครบทุกตัวอักษร แต่ต้องปรับให้คำที่เกิดจากการถ่ายถอดอักษร ออกเสียงได้

ใกล้เคียงกับคำดั้งเดิมมากที่สุด การถ่ายถอดอักษรประเภทนี้จะคำนึงถึงคุณสมบัติทางเสียงมากกว่าตัวอักษร ลักษณะเช่นนี้จึงคล้ายกับการถ่ายถอดอักษรแบบถ่ายเสียง ตัวอย่างเช่น คำว่า “สุวรรณภูมิ” เมื่อทำการถ่ายถอดอักษรตามแนวทางนี้จะได้เป็น Suwannaphum เป็นต้น

เป็นที่น่าสังเกตว่า ข้อดีของการถ่ายถอดอักษรแบบยืมเสียงเป็นสำคัญคือ ทำให้คำที่ถ่ายถอดอักษรมีเสียงใกล้เคียงกับคำดั้งเดิมซึ่งจะก่อให้เกิดประโยชน์ต่าง ๆ เช่น การเรียนภาษาต่างประเทศ แต่การถ่ายถอดอักษรวิธีการนี้ จะส่งผลให้การค้นคืนข้ามภาษาทำได้ยาก เนื่องจากกฎในการออกเสียงของแต่ละภาษามีความแตกต่างกัน การจะถ่ายถอดอักษรได้ถูกต้องนั้น ต้องอาศัยกฎที่จำเป็นสำหรับการถ่ายถอดอักษรจากภาษาหนึ่งไปยังอีกภาษาหนึ่งที่ยังพอทำให้กระบวนการในการทำงานเกิดความซับซ้อนมากยิ่งขึ้น

อย่างไรก็ตาม หากพิจารณาทิศทาง (direction) ในการถ่ายถอดอักษรระหว่างภาษาต้นทางและภาษาปลายทาง ก็สามารถแบ่งออกได้เป็น 2 ประเภท คือการถ่ายถอดอักษรแบบปกติและการถ่ายถอดอักษรแบบย้อนกลับ (Lee, C.-J. and Chang, J. S. 2003) การถ่ายถอดอักษรแบบปกติหรือแบบไปข้างหน้า (forward-direction transliteration) เป็นกระบวนการในการแปลงคำดั้งเดิม (original word) ในภาษาต้นทางให้เป็นคำใหม่ในภาษาปลายทางโดยให้มีเสียงใกล้เคียงกับภาษาต้นทาง เช่นคำว่า “ต้มยำกุ้ง” ซึ่งเป็นคำเดิมในภาษาไทยสามารถถอดอักษรได้เป็น “tom yum kung” ในภาษาอังกฤษ อีกประเภทหนึ่งคือการถ่ายถอดอักษรแบบย้อนกลับ (backward-direction transliteration) ซึ่งเป็นการถ่ายถอดอักษรจากคำยืมหรือคำทับศัพท์จากภาษาหนึ่ง กลับเป็นตัว อักษรดั้งเดิมในอีกภาษาหนึ่งเช่น คำว่า “tom yum kung” ซึ่งเป็นคำภาษาอังกฤษที่เกิดจากการถ่ายถอดมาจากภาษาไทยถูกแปลงกลับให้เป็น “ต้มยำกุ้ง” ซึ่งเป็นคำดั้งเดิมที่ปรากฏในภาษาไทยอยู่แล้ว

### 2.3.2 แนวทางการถ่ายถอดอักษรด้วยคอมพิวเตอร์

Oh, Choi และ Isahara (2006: 186) ได้แบ่งวิธีการในการถ่ายถอดอักษรด้วยคอมพิวเตอร์ออกเป็น 3 ประเภท ได้แก่ (1) การถ่ายถอดอักษรโดยการเทียบตัวอักษร (grapheme based approach หรือ direct method) (2) การถ่ายถอดอักษรโดยอาศัยตัวกลางทางเสียง (phoneme based approach หรือ pivot method) และ (3) การถ่ายถอดอักษรแบบผสมผสาน (hybrid approach) ซึ่งเป็นการผสมผสานระหว่างการอาศัยตัวกลางทางเสียงและการเทียบตัวอักษรเข้าด้วยกัน ซึ่งจะขอนำเสนอรายละเอียดดังนี้



### 2.3.2.1 การถ่ายถอดอักษร โดยการเทียบตัวอักษร

การถ่ายถอดอักษรโดยการเทียบตัวอักษร (Grapheme based approach) เป็นการจับคู่ระหว่างตัวเขียนภาษาต้นทางกับตัวเขียนของภาษาปลายทางโดยไม่ผ่านตัวกลางทางเสียง ซึ่งตัวอักษรที่จับคู่นี้อาจเป็นความสัมพันธ์แบบหนึ่งต่อหนึ่งหรือไม่ใช่หนึ่งต่อหนึ่งก็ได้ (Oh, J.-H. Choi, K.-S. and Isahara, H. 2006: 188)

ในการถ่ายถอดอักษรแบบเทียบตัวอักษรสามารถใช้วิธีการต่างๆ ได้แก่ การใช้กฎทางภาษา (Rule-based model) และ การใช้แบบจำลองทางสถิติ (Statistical machine transliteration model หรือ SMT) ซึ่งมีรายละเอียดดังนี้

#### 2.3.2.1.1 การใช้กฎทางภาษา

การใช้กฎทางภาษา (Rule-based model) คือ การนำกฎทางภาษามาใช้ในการถ่ายถอดอักษรจากภาษาหนึ่งไปสู่อีกภาษาหนึ่ง ซึ่งกฎต่าง ๆ ในการถ่ายถอดอักษรอาจได้มาจาก การนำกฎที่มนุษย์สร้างขึ้นไปใช้กับเครื่องคอมพิวเตอร์โดยตรง (manual construction) หรือการหากฎด้วยคอมพิวเตอร์โดยอาศัยคลังข้อความ (corpus-based) (วิโรจน์ อรุณมานะกุล 2548:37)

สำหรับกฎที่กำหนดขึ้น โดยมนุษย์นั้น อาจจำแนกได้เป็น (1) กฎที่มีผู้กำหนดไว้แล้วเช่นกฎที่ปรากฏในตำราต่าง ๆ รวมทั้งกฎที่ตราขึ้นเป็นมาตรฐานโดยภาครัฐ เช่น เกณฑ์ศัพท์บัญญัติของราชบัณฑิตยสถาน หรือ (2) กฎที่เกิดจากการศึกษาเพิ่มเติมในภายหลัง ไม่ว่าจะเป็นกฎที่ได้จากผู้เชี่ยวชาญ หรือการทำงานวิจัยเพิ่มเติม ฯลฯ ส่วนการหากฎด้วยคอมพิวเตอร์โดยอาศัยคลังข้อความนั้นทำได้โดยเก็บรวบรวมคลังข้อความ (corpus) ที่ต้องการจะหากฎให้ได้จำนวนมากพอสมควร จากนั้นเขียน โปรแกรมให้คอมพิวเตอร์ได้เรียนรู้จากคลังข้อความนั้นๆภายใต้บริบทต่างๆ แล้วนำกฎที่เครื่องคอมพิวเตอร์หาได้มาใช้ในการประมวลผลต่อไป

ในการถ่ายถอดอักษรด้วยเครื่องคอมพิวเตอร์โดยใช้กฎทางภาษา อาจทำได้โดย การใช้ผังต้นไม้ช่วยตัดสินใจ (decision trees) การใช้ตารางกฎ (table lookup) และการจับคู่อักษร (character mapping) เป็นต้น

#### ก. ผังต้นไม้ช่วยตัดสินใจ

โดยทั่วไป ในการใช้ผังต้นไม้ช่วยตัดสินใจ (decision trees) จะมีการกำหนดผังต้นไม้ซึ่งมีลักษณะคล้ายต้นไม้จริงกลับหัวที่มีโหนดรากอยู่ด้านบนบนสุดและโหนดใบอยู่ด้านล่างสุด โดยที่แต่ละกิ่งแทนแต่ละขั้นตอนของการตัดสินใจ ในการประมวลผลจะเริ่มจากด้านบนบนสุดของผังต้นไม้แล้วไล่ไปตามกิ่งต่าง ๆ เป็นลำดับ ตามเงื่อนไขที่ระบุไว้ในแต่ละกิ่งจนสิ้นสุดผังต้นไม้แล้วให้ผลลัพธ์สุดท้ายเป็นผลลัพธ์ในการตัดสินใจ ในการถ่ายถอดโดยวิธีนี้จึงเป็นการ

พิจารณาตัวอักษรพร้อมกับบริบทข้างเคียงเพื่อตัดสินใจที่ละขั้นตามลำดับจนถึงขั้นสุดท้ายคือ ปลายกิ่งจนได้ตัวอักษรเป้าหมาย (สุชชาติรี ประสมสุข. 2555)

กั้งและชอย (Kang, I.-H. and Choi, K.-S. 2000) ได้เสนอแบบจำลองการถ่ายทอดจาก ภาษาอังกฤษเป็นภาษาเกาหลีโดยการใช้ผังต้นไม้ช่วยตัดสินใจ โดยที่ในขั้นแรก กั้งและชอยได้ทำ การปรับแนว (align) ตัวอักษรของข้อความในภาษาต้นทางกับตัวอักษรในภาษาปลายทาง แล้วจึง ใช้ผังต้นไม้มาช่วยตัดสินใจว่าสายอักขระใดเหมาะสมที่สุดที่จะนำมาใช้เป็นผลลัพธ์ของการถ่าย โยดอักษรในภาษาปลายทางได้

#### ข. การใช้ตารางกฎ

การใช้ตารางกฎ (table lookup) เป็นวิธีการที่ต้องอาศัยตารางที่ประกอบด้วยกฎต่าง ๆ ที่ สร้างขึ้น โดยที่ตารางกฎนี้จะประกอบด้วยตัวอักษรแต่ละชุด (โดยที่แต่ละชุดจะประกอบด้วยอักษร 1 ตัว 2 ตัวหรือ 3 ตัวก็ได้ แล้วแต่กรณี) โดยแต่ละชุดอักษรนี้ จะมีการระบุถึงตัวอักษรทางขวาและ ตัวอักษรทางซ้ายที่ปรากฏร่วมด้วย และระบุว่าหากพบบริบทตัวอักษรทางซ้ายและทางขวานี้แล้วจะ ถ่ายทอดอักษรเป็นตัวอักษรใดในภาษาปลายทาง

ในการถ่ายทอดอักษร คอมพิวเตอร์จะนำตัวอักษรแต่ละชุดพร้อมกับบริบทซ้าย-ขวาของคำ ที่ต้องการถ่ายทอดมาเปรียบเทียบกับชุดอักษรในตารางทีละชุด เพื่อตรวจสอบว่าชุดอักษรและ บริบทซ้ายและขวาของคำที่ต้องการถ่ายโอนนั้นตรงกับชุดอักษรและบริบทใดในตาราง แล้วนำ ผลลัพธ์ของกรณีนั้นมาใช้เป็นผลลัพธ์ของการถอดชุดอักษรนั้น แต่หากพบว่าชุดอักษรนั้นไม่ตรง กับกรณีใด หรือพบว่าชุดอักษรนั้นตรงกับกรณีต่าง ๆ มากกว่า 1 กรณีในตาราง คอมพิวเตอร์ก็จะทำ การตรวจสอบกับคลังข้อมูลฝึกอบรมที่ได้เตรียมไว้แล้ว เพื่อหาว่าชุดอักษรที่มีปัญหานี้มีโอกาสที่จะ ถอดเป็นอักษรใดได้มากที่สุด แล้วจึงสร้างตารางคาดเดาขึ้นเพื่อนำมาใช้ในการประมวลผลชุดอักษร ที่มีปัญหานั้น (วิโรจน์ อรุณมานะกุล 2548)

วัลย์วรา ไชยฤกษ์ (2547) ได้ศึกษาการถ่ายทอดคำทับศัพท์ภาษาอังกฤษเป็นภาษาไทย โดย เริ่มจากสร้างตารางกฎซึ่งได้มาจากข้อกำหนดของราชบัณฑิตยสถาน โดยตารางกฎที่ใช้มี 2 ประเภท ได้แก่ ตารางกฎที่มีบริบทซ้าย-ขวาอย่างละ 1 ตำแหน่ง 2 ตำแหน่งและ 3 ตำแหน่ง และตารางคาดเดา ซึ่งประกอบด้วยตารางคาดเดาที่มีบริบทซ้าย-ขวา 2 ตำแหน่ง 1 ตำแหน่ง และตารางคาดเดาที่ไม่มีใน บริบท

#### ค. Character Mapping

Character Mapping เป็นการจับคู่ตัวอักษรของภาษาหนึ่งกับอักษรของอีกภาษาหนึ่ง โดยตรง (Min, Z. Haizhou, L. and Jian, S. 2004)

Deep, Kamal and Goyal, Vishal (2011) ได้ถ่ายถอดอักษรชื่อบุคคลและชื่อสถานที่ที่ ภูมิศาสตร์จากภาษาปัญจาบ (Punjabi) เป็นภาษาอังกฤษ โดยใช้ Character Mapping ในการจับคู่ ตัวอักษรระหว่าง Gurumakhi script กับ English letter โดยในการถ่ายถอดอักษรได้จำแนกกฎ ออกเป็น 2 ประเภท คือ กฎที่ใช้ในการจับคู่ตัวอักษรระหว่างภาษาทั้งสอง และ กฎเกี่ยวกับตำแหน่ง ของการปรากฏของตัวอักษรของคำในภาษาปัญจาบ เช่นปรากฏที่ตำแหน่งต้นคำ ท้ายคำ หลังสระ หรือหลังพยัญชนะ เป็นต้น

#### 2.3.2.1.2 การใช้แบบจำลองทางสถิติ

การถ่ายถอดอักษรโดยใช้แบบจำลองทางสถิติ (Statistical machine transliteration model หรือ SMT) เป็น แบบจำลองที่ใช้ข้อมูลสถิติที่พบจากการถ่ายถอดอักษรจริงมาช่วยในการพัฒนา ระบบ แบบจำลองนี้มองว่า การถ่ายถอดอักษรมีความเป็นไปได้ที่จะให้ผลลัพธ์ในหลาย ๆ รูปแบบ โดยที่แบบจำลองทางสถิติจะเป็นตัวเลือกรูปแบบที่เหมาะสมมากที่สุดมาใช้เป็นผลลัพธ์

การพัฒนาระบบที่ใช้แบบจำลองทางสถิติ กระทำได้ในหลายลักษณะ ดังนี้

##### ก. Noisy Channel Model

แบบจำลองนี้ใช้ในการปรับข้อมูลนำเข้าให้เป็นผลลัพธ์ที่ต้องการ เนื่องจากอิงอยู่กับ สมมติฐานว่าข้อมูลที่นำเข้ามานี้มีข้อผิดพลาด จึงต้องให้คอมพิวเตอร์ทำกระบวนการต่าง ๆ เพื่อให้ได้ ผลลัพธ์เป็นข้อมูลที่ถูกต้อง เช่น หากใช้แบบจำลองนี้มาแก้คำผิดก็จะเชื่อว่าคำที่นำเข้าสะกดไม่ ถูกต้อง เนื่องจากอาจมีอักขระบางตัวหายไป (เช่น letter แทนที่จะเป็น letter) มีอักขระเกินมา (เช่น mistake แทนที่จะเป็น mistake) มีการสลับที่อักขระ (เช่น receiving แทนที่จะเป็น receiving) หรือ มีการแทนที่อักขระตัวหนึ่งด้วยอักขระอีกตัวหนึ่ง (เช่น finite แทนที่จะเป็น finite) เป็นต้น

ในแบบจำลองนี้ประกอบด้วยแบบจำลองย่อยอีก 2 แบบจำลองคือ (1) Noisy Channel model เป็นแบบจำลองที่จะสร้างเซตของคำเป้าหมายที่เป็นไปได้ทั้งหมดจากคำนำเข้า และ (2) Language model เป็นแบบจำลองที่หาความเป็นไปได้สูงสุดจากเซตของคำที่เครื่องคอมพิวเตอร์สร้างขึ้นจาก แบบจำลองที่ผ่านมา โดยที่รูปแบบที่ใช้อาจเป็น maximum likelihood, maximum a posteriori หรือ minimum distance เป็นต้น

Jia, Zhu and Yu (2009) ได้ใช้ noisy channel model ในการถ่ายถอดภาษาอังกฤษเป็น ภาษาจีนแบบไปข้างหน้าและย้อนกลับเพื่อใช้ในงานค้นคืนข้ามภาษาแนวคิดคือการมองว่าจากคำ ภาษาอังกฤษที่กำหนดให้ (E) ซึ่งอาจเป็นคำจีน (C) ได้หลายแบบ โดยขั้นแรกเมื่อนำเข้าข้อมูลที่เป็นคำ ภาษาอังกฤษแบบจำลองนี้จะสร้างคำจีน (C) ที่เป็นไปได้ทั้งหมดขึ้นมา จากนั้นใช้กฎของ Bayes Rule เพื่อจะหาคำจีน ( $\hat{C}$ ) ซึ่งเป็นคำที่สอดคล้องกับคำอังกฤษมากที่สุด

ข. ระบบที่ใช้แบบจำลองจากคู่ตัวอักษรสองภาษา (Joint source-channel model)

ระบบดังกล่าวจะใช้ข้อมูลจากทั้งภาษาต้นทางและภาษาปลายทางไปพร้อมๆกัน ระบบนี้มีแนวคิดหลักที่ว่า ในการตัดสินใจเลือกตัวอักษรตัวต่อไปในภาษาปลายทางนั้น ขึ้นอยู่กับข้อมูลของทั้งในภาษาต้นทางและในภาษาปลายทางที่มาก่อนหน้า วิธีการนี้มีข้อดีคือ ไม่จำเป็นต้องใช้ข้อมูลคำอ่านของทั้งสองภาษา เนื่องจากบริบทของทั้งสองภาษาจะช่วยให้เลือกตัวอักษรในภาษาปลายทางที่เหมาะสมได้อยู่แล้ว

วิโรจน์ อรุณมานะกุล (2548: 37-38) ได้นำวิธีการนี้มาใช้เป็นส่วนหนึ่งในการถ่ายถอดคำภาษาอังกฤษเป็นภาษาไทย โดยใช้โปรแกรมในการดูค่าความน่าจะเป็นสูงสุดของการปรากฏของตัวอักษรไทยและอังกฤษไปพร้อม ๆ กัน ในการถ่ายถอดอักษรนั้นจะยึดหลักการที่ว่า คำอังกฤษหนึ่ง คำ สามารถถ่ายถอดให้เป็นคำไทยได้หลายรูปแบบ ส่วนรูปแบบของคำไทยที่เหมาะสมที่สุดนั้นขึ้นอยู่กับปรากฏของตัวอักษร 3 ตัวในภาษาอังกฤษที่ปรากฏร่วมกันตั้งแต่ต้นคำถึงท้ายคำ

ค. ระบบที่ใช้แบบจำลองจากคู่ตัวอักษรสองภาษาและมองเป็นกลุ่มก้อนได้ (Chunk-based)

วิธีการนี้มีแนวคิดว่าการถ่ายถอดอักษรไม่จำเป็นต้องเทียบคู่ตัวอักษรระหว่างภาษาต้นทางและภาษาปลายทาง แต่มองความสัมพันธ์ระหว่างตัวอักษรของภาษาต้นทางและภาษาปลายทางนั้นเป็นแบบกลุ่มตัวอักษรต่อกัน โดยที่กลุ่มของตัวอักษรสามารถมีความยาวตั้งแต่หนึ่งตัวอักษรขึ้นไปได้

ตัวอย่างเช่น locarno = ล/โ/ค/า/ร/โน นอกจากจะมองว่า 'l' ถ่ายถอดเป็น 'ล', 'o' ถ่ายถอดเป็น 'โ', 'c' ถ่ายถอดเป็น 'ค' แล้ว ยังสามารถมองการถ่ายถอดอักษรเป็นกลุ่มของตัวอักษรที่มีขนาดความยาวต่างกันไม่ว่าจะเป็น 2, 3, 4 จนถึง n ตัวอักษร เช่น 'lo' เป็น 'ลโ', 'oc' เป็น 'โค', 'ca' เป็น 'คา', 'ar' เป็น 'าร์', no เป็น 'โน', 'loc' เป็น 'ลโค' 'locarno' เป็น 'ลโคอาร์โน' เป็นต้น (วิโรจน์ อรุณมานะกุล 2548: 39) ดังนั้น จากข้อมูลฝึกสอนระบบซึ่งได้จับคู่ตัวอักษรระหว่างภาษาต้นทางและภาษาปลายทางแล้ว โปรแกรมจะคำนวณค่าสถิติของการถ่ายถอดกลุ่มอักษรต่างๆ หลังจากนั้นจะนำค่าสถิติที่ได้ไปใช้ในการตัดสินใจถ่ายถอดอักษรอื่นๆ ต่อไป

เท่าที่ได้อภิปรายมาพอจะกล่าวได้ว่า การถ่ายถอดอักษรโดยการเทียบตัวอักษร เป็นระบบที่ง่ายต่อการพัฒนา ระบบนี้มีข้อดีคือใช้เพียงข้อมูลของตัวเขียนทั้งสองภาษาซึ่งเป็นข้อมูลพื้นฐานที่จำเป็นต้องจัดเตรียมไว้แล้ว แต่การถ่ายถอดระหว่างตัวเขียนของภาษาหนึ่งไปยังอีกภาษาหนึ่งโดยตรงมีข้อจำกัดคืออาจได้ผลไม่ถูกต้องในกรณีที่ตัวเขียนและเสียงในภาษานั้นมีความสัมพันธ์ที่ซับซ้อน ซึ่งในกรณีนี้ควรเลือกใช้การถ่ายถอดโดยอาศัยตัวกลางทางเสียงซึ่งจะได้อภิปรายในหัวข้อต่อไป

### 2.3.2.2 การถ่ายถอดอักษร โดยอาศัยตัวกลางทางเสียง

การถ่ายถอดโดยอาศัยตัวกลางทางเสียง (phoneme based approach) เป็นการสะกดคำในภาษาปลายทางที่ทำให้ออกเสียงใกล้เคียงกับเสียงพูดของคำดั้งเดิมในภาษาต้นทางมากที่สุด

วิโรจน์ อรุณมานะกุล (2548: 30) กล่าวว่า การถ่ายถอดอักษรโดยใช้เสียงเป็นตัวกลางสามารถทำได้หลายลักษณะคือ

ก. รูปแบบที่ถ่ายอักษรภาษาต้นทางเป็นเสียงภาษาต้นทางแล้วเทียบเป็นเสียงในภาษาปลายทาง ก่อนจะถอดออกมาเป็นตัวอักษรในภาษาปลายทาง (Source Language Texts->Source Language Sound ->Target Language Sound-> Target Language Texts) การถ่ายถอดอักษรรูปแบบนี้ ทำโดยผ่านตัวกลางคือเสียงพูดของทั้งภาษาต้นทางและปลายทาง ดังที่ปรากฏในงานวิจัยของ ลี (Lee, J. S. 1999) ไนท์และเกรล (Knight, K. and Graehl, J. 1997)

ข. รูปแบบที่ถ่ายเสียงตัวอักษรภาษาต้นทางเป็นเสียงในภาษาต้นทางแล้วแปลงเป็นตัวอักษรในภาษาปลายทางโดยลดขั้นตอนการเทียบหน่วยเสียงจากภาษาต้นทางเป็นหน่วยเสียงภาษาปลายทาง (Source Language Texts -> Source Language Sounds -> Target Language Texts ) ดังปรากฏในงานของ Glover and Knight (1998) ที่ถ่ายถอดจากภาษาอังกฤษเป็นอารบิก

ค. รูปแบบที่ถ่ายเสียงตัวอักษรภาษาต้นทางไปเป็นเสียงในภาษาปลายทางโดยตรงก่อนจะถอดเป็นตัวอักษรในภาษาปลายทาง (Source Language Texts -> Target Language Sounds -> Target Language Texts ) ดังปรากฏในงานของ Kuo and Yang (2004) ที่ถ่ายถอดจากภาษาอังกฤษเป็นภาษาจีน

ในการถ่ายถอดอักษร โดยอาศัยตัวกลางทางเสียง สามารถทำได้โดย การใช้กฎทางภาษา การใช้แบบจำลองทางสถิติ การใช้กฎร่วมกับแบบจำลองทางสถิติ และ การใช้เครือข่ายประสาทเทียม ซึ่งมีรายละเอียดดังนี้

#### 2.3.2.2.1 การใช้กฎทางภาษา

วิธีการใช้กฎทางภาษาทำได้หลายวิธี วิธีการหนึ่งคือการใช้รหัสตัวเลขหรือที่เรียกว่า Soundex Technique ซึ่งริเริ่มโดย M. K. Odell และ R. C. Russell ในปี พ.ศ.2461

ชาวค็เด็คส์คือรหัสทางเสียง (phonetic code) ที่ได้มาจากการเปลี่ยนตัวอักษรต่าง ๆ ให้เป็นตัวเลข (numeric code) โดยมีหลักการคือ ต้องมีการเก็บรวบรวมข้อมูลคำในภาษาต้นทางและภาษาปลายทาง แล้วนำคำในภาษาทั้งสองที่อยู่ในคลังข้อมูลมาเปลี่ยนเป็นรหัสชาวค็เด็คส์ของแต่ละภาษา หลังจากนั้นเมื่อต้องการถอดคำในภาษาใดภาษาหนึ่งก็จะนำคำนั้นมาเปลี่ยนเป็นรหัสชาวค็เด็คส์

แล้วนำรหัสนี้ไปใช้เป็น key ในการค้นว่าตรงกับรหัสชาวด์เด็กซ์ของอีกภาษาหนึ่งหรือไม่ ถ้าตรงกันก็ให้นำคำที่ตรงกับรหัสชาวด์เด็กซ์ในภาษาปลายทางมาเป็นผลลัพธ์

Suwanvisat and Prasitjutrakul (1998) ได้นำวิธีการนี้มาใช้ในการถ่ายถอดอักษรระหว่างภาษาไทยและอังกฤษ เพื่อใช้ในการสืบค้นเอกสารอิเล็กทรอนิกส์ (retrieval of document) โดยได้ปรับตารางรหัสและอัลกอริทึมของโอเคิลและรัสเซลเพื่อให้คอมพิวเตอร์ประมวลผลได้ดีขึ้น

### 2.3.2.2.2 การใช้แบบจำลองทางสถิติ

การถ่ายถอดอักษรที่อาศัยตัวกลางทางเสียงโดยใช้แบบจำลองทางสถิติ สามารถกระทำได้ในหลายลักษณะ ดังนี้

#### ก. การใช้แบบจำลองเอ็นแกรม

แบบจำลองเอ็นแกรม (N-gram model) คือแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของสายอักขระที่เรียงต่อกันจำนวน  $n$  ตัว แบบจำลองเอ็นแกรมจะตั้งอยู่บนสมมติฐานมาร์คอฟ (Markov assumption) ที่ว่า “การปรากฏของตัวอักษรตัวหนึ่งขึ้นกับตัวอักษรก่อนหน้าเพียง  $n-1$  ตัว” นอกจากนี้เราสามารถนำสมมติฐานมาร์คอฟมาใช้ประมาณค่าความเป็นไปได้ที่ 2-gram 3-gram 4-gram ฯลฯ ได้

ลี (Lee, J. S. 1999) ได้ศึกษาการถ่ายถอดภาษาอังกฤษเป็นภาษาเกาหลีโดยใช้แบบจำลองเอ็นแกรมเพื่อหาความน่าจะเป็นสูงสุดของชุดตัวอักษรต่าง ๆ ในภาษาเกาหลีที่ได้มาจากคำภาษาอังกฤษ ในการถ่ายถอด ลีจะใช้การจับคู่ระหว่างหน่วยเสียงพูดของภาษาอังกฤษและภาษาเกาหลี วิธีการเริ่มจากนำคำอังกฤษมาแยกออกเป็นหน่วยเสียงพูด (PU's) เช่น 'board' อาจแบ่งได้เป็น b:o:a:r:d, b:oa:r:d, b:o:ar:d b:oar:d ฯลฯ (ซึ่ง ลี ใช้สัญลักษณ์แทนแต่ละรูปแบบด้วย  $E_i$ ) หลังจากนั้นนำแต่ละ  $E_i$  ไปสร้างรูปแบบหน่วยเสียงพูดในภาษาเกาหลีให้ได้จำนวนมากที่สุด แล้วคำนวณหารูปแบบหน่วยเสียงพูดในภาษาเกาหลีที่มีความเป็นไปได้สูงสุด แล้วนำรูปแบบที่ได้นี้ไปจับคู่กับตัวเขียนในภาษาเกาหลีเพื่อใช้เป็นผลลัพธ์

ค่าความน่าจะเป็น  $P(K)$  ถูกกำหนดโดยโมเดลของภาษาและสามารถประมาณได้โดยการใช้แบบจำลองไบแกรม ในการคำนวณหาหน่วยเสียงพูดที่เป็นไปได้สูงสุดในภาษาเกาหลีทำได้โดยสมการดังนี้

$$\arg \max_k P(K | E) = \arg \max_k P(K)P(E | K)$$

$$P(K) \cong p(kpu_1) \prod_{i=2}^n p(kpu_i | kpu_{i-1})$$

$$P(K | E) \cong \prod_{i=1}^n p(epu_i | kpu_i)$$