การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน

นางสาววัลย์วรา ไชยฤกษ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรมหาบัณฑิต
สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2547
ISBN 974-17-6489-8
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

THE DEVELOPMENT OF AN ENGLISH-THAI TRANSLITERATION PROGRAM BASED ON TRANSLITERATED-WORD CORPUS OF THE ROYAL INSTITUTE

Miss Wanwara Chairoek

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Linguistics

Department of Linguistics
Faculty of Arts
Chulalongkorn University

Academic Year 2004
ISBN 974-17-6489-8

หัวข้อวิทยานิพนธ์	การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน
โดย	นางสาววัลย์วรา ไชยฤกษ์
ภาควิชา	ภาษาศาสตร์
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล

คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะอักษรศาสตร์
(ศาสตราจารย์ ดร.ธีระพันธ์ เหลืองทองคำ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุดาพร ลักษณียนาวิน)

..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล)

..... กรรมการ
(รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วัลย์วรา ไชยฤกษ์ : การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน. (THE DEVELOPMENT OF AN ENGLISH-THAI TRANSLITERATION PROGRAM BASED ON TRANSLITERATED-WORD CORPUS OF THE ROYAL INSTITUTE) อ. ที่ปรึกษา: ผู้ช่วยศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล, 150หน้า. ISBN 974-17-6489-8.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อรวบรวมคำทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถาน และเปรียบเทียบกฎการทับศัพท์ของราชบัณฑิตยสถานกับข้อมูลจริงที่พบ พร้อมทั้งสร้างโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ดังกล่าวเพื่อสร้างตารางกฎการถอดอักษรภาษาอังกฤษเป็นภาษาไทยให้กับคอมพิวเตอร์

การวิจัยนี้ได้เก็บรวบรวมคำภาษาอังกฤษและคำทับศัพท์ภาษาไทยจำนวน 10,060 คำจากเอกสารที่เผยแพร่โดยราชบัณฑิตยสถานเพื่อนำมาใช้เป็นคลังคำทับศัพท์ ผลการวิจัยพบว่า กฎการถอดอักษรของราชบัณฑิตยสถานสอดคล้องกับข้อมูลในคลังทับศัพท์ดังนี้ กฎสำหรับสระมีความสอดคล้อง 92.68% สำหรับพยัญชนะต้นสอดคล้อง 97.35% สำหรับพยัญชนะท้ายสอดคล้อง 97.34% และเมื่อ ทดลองโปรแกรมที่พัฒนาขึ้นโดยใช้กฎที่สร้างขึ้นจากข้อมูล 80% ของคลังคำทับศัพท์ ผลการทดลองกับข้อมูลอีก 20% ปรากฏว่าโปรแกรมสามารถถอดอักษรในระดับคำได้ถูกต้อง 39.53% และถูกต้อง 79.8% ในระดับอักษร ซึ่งแสดงว่าประสิทธิภาพการถอดอักษรของโปรแกรมยังไม่ดีเท่าที่ควร เนื่องมาจากสาเหตุหลายประการ เช่น โปรแกรมไม่ได้นำขอบเขตพยางค์มาพิจารณา ไม่ได้พิจารณาเสียงภาษาอังกฤษ รวมถึงความหลากหลายของคำทับศัพท์ของราชบัณฑิตยสถานที่รวบรวมไว้ในคลังข้อมูล และยังพบว่าในการถอดอักษรจริงนั้น มีการถอดอักษรแบบที่นอกเหนือไปจากที่ราชบัณฑิตยสถานให้ไว้ในตารางเทียบอักษรอยู่จำนวนหนึ่ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....ภาษาศาสตร์..... ลายมือชื่อนิสิต.....
สาขาวิชา.....ภาษาศาสตร์..... ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2547

4480195322 : MAJOR LINGUISTICS

KEY WORD: TRANSLITERATION / CORPUS / ENGLISH / THAI / THE ROYAL INSTITUE

WANWARA CHAIROEK : THE DEVELOPMENT OF AN ENGLISH-THAI TRANSLITERATION PROGRAM BASED ON TRANSLITERATED-WORD CORPUS OF THE ROYAL INSTITUTE. THESIS ADVISOR: WIROTE AROONMANAKUN, Ph.D., 150 pp. ISBN 974-17-6489-8.

The purpose of this thesis is to collect transliterated words published by the Royal Institute and compare the transliteration regulation of the Royal Institute with the collected data. Furthermore, an English-Thai transliteration program based on transliterated-word corpus of the Royal Institute was developed in order to generate tables of English-Thai transliteration rules for computers.

The transliterated-word corpus is composed of 10,060 pairs of English and transliterated Thai words published by the Royal Institute. The result shows that the transliteration regulation of the Royal institute conforms with the collected data in the corpus as follows: The transliteration rules for vowels have 92.68% accordance, the transliteration rules for initial consonants have 97.35% accordance and the transliteration rules for last consonants have 97.34% accordance. When the developed program was trained by the rules generated from 80% of the corpus, the result of testing on 20% remainder shows 39.53% accuracy in word level and 79.8% accuracy in alphabetical level. This suggests that the transliteration program yields quite a poor result due to considerable reasons, for instance syllable boundaries and phonetic transcriptions were not considered including various versions of transliterated words published by the Royal Institute collected in the corpus. It is also found that a number of transliteration rules are not yet included in the Royal Institute's table of transliteration.

Department.....Linguiatics..... Student's signature.....

Field of study...Linguistics..... Advisor's signature.....

Academic year 2004

กิตติกรรมประกาศ

ผู้วิจัยต้องขอขอบพระคุณ ผ.ศ.ดร.วิโรจน์ อรุณมานะกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างสูง วิทยานิพนธ์เล่มนี้คงจะสำเร็จลุล่วงไปด้วยดีไม่ได้หากไม่ได้รับความช่วยเหลือและคำปรึกษาต่างๆเกี่ยวกับวิทยานิพนธ์เล่มนี้จากอาจารย์ และผู้วิจัยขอขอบพระคุณ ผ.ศ.ดร.สุดาพร ลักษณียนาวิน และ ร.ศ.ดร.บุญเสริม กิจศิริกุล กรรมการสอบวิทยานิพนธ์ที่ได้ให้คำปรึกษาและเสียสละเวลาเพื่อตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้

นอกจากนี้ผู้วิจัยขอขอบคุณคุณมานะ จันทร์สมบูรณ์ สำหรับคำชี้แนะแนวทางในการทำวิทยานิพนธ์ ขอขอบคุณคุณมธุรส จิระพรพงศ์และคุณอนนต์ศม สุขสมัย เพื่อนผู้คอยกระตุ้นเตือนและให้กำลังใจผู้วิจัยมาตลอด และขอขอบคุณคณาจารย์ภาควิชาภาษาศาสตร์ทุกท่านที่ได้ประสิทธิ์ประสาทความรู้ด้านภาษาศาสตร์ให้แก่ผู้วิจัย รวมทั้งขอขอบคุณเจ้าหน้าที่และเพื่อนพี่น้องภาควิชาภาษาศาสตร์ทุกคนที่ช่วยเหลือและให้กำลังใจผู้วิจัยเสมอมา

สุดท้ายนี้ผู้วิจัยขอขอบคุณคุณอิสรศักดิ์ ติสถาพร เป็นอย่างยิ่ง ที่ได้ช่วยเหลือและให้กำลังใจในการทำวิทยานิพนธ์อย่างดียิ่งเสมอมา ขอขอบคุณคุณวรัชญ์ ไชยฤกษ์ น้องชายที่ช่วยอำนวยความสะดวกในการทำวิทยานิพนธ์ด้วยดีมาตลอดและผู้วิจัยต้องขอขอบพระคุณบุคคลสำคัญที่ขาดไม่ได้คือบิดามารดาคุณวราตุลย์และคุณสภัสสร ไชยฤกษ์ที่ได้เลี้ยงดูอบรมสั่งสอนผู้วิจัยมาเป็นอย่างดีจนถึงวันนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อวิทยานิพนธ์ภาษาไทย ..... ง
บทคัดย่อวิทยานิพนธ์ภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ฎ

บทที่

1. บทนำ 1
    1.1 ความเป็นมาและความสำคัญของปัญหา 1
    1.2 วัตถุประสงค์ของการวิจัย 3
    1.3 สมมติฐานของการวิจัย 3
    1.4 กรอบทฤษฎี 4
    1.5 เครื่องมือที่ใช้ในงานวิจัย 4
    1.6 ขอบเขตของการวิจัย 4
    1.7 ประโยชน์ที่คาดว่าจะได้รับ 4
    1.8 วิธีดำเนินการวิจัย 4
    1.9 โครงสร้างของวิทยานิพนธ์ 5
2. เอกสารและงานวิจัยที่เกี่ยวข้อง 7
    2.1 ประวัติความเป็นมาของการถอดอักษรในภาษาไทย 7
    2.2 คำจำกัดความความแตกต่างและลักษณะเฉพาะของการถอดอักษรและการถ่ายเสียง 11
    2.3 ปัญหาทางภาษาศาสตร์และปัญหาอื่นๆในการถอดอักษร 15
    2.4 หลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถาน 17
    2.5 แนวทางในการแปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่ง 35
    2.6 แนวทางในการวางกฎในการถอดอักษรให้คอมพิวเตอร์เรียนรู้ 38
    2.7 การจับคู่ข้อมูล 39
    2.8 แบบจำลองที่ใช้ 40

หน้า

บทที่

3. วิธีดำเนินการวิจัย 43

3.1 การเก็บรวบรวมข้อมูล 45

3.2 การจับคู่ข้อมูลภาษาอังกฤษและภาษาไทย 47

3.3 การเขียนโปรแกรมนำข้อมูลใส่ตารางตามบริบท 50

3.4 การเปรียบเทียบความสอดคล้องของกฎการถอดอักษรที่ได้จากคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน 54

3.5 การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน 54

3.6 การประเมินผลความถูกต้องแม่นยำของโปรแกรมการถอดอักษร 56

4. ผลการถอดอักษรโดยใช้คอมพิวเตอร์ 58

4.1 ผลการเปรียบเทียบความสอดคล้องของกฎที่ได้จากคอมพิวเตอร์และกฎของราชบัณฑิตยสถาน 58

4.2 ผลการทดลองถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน 75

5. สรุปและข้อเสนอแนะ 85

5.1 สรุปกระบวนการในการพัฒนาโปรแกรมการถอดอักษรและผลการทำงานของโปรแกรมการถอดอักษร 85

5.2 ข้อเสนอแนะในการศึกษาพัฒนาเพิ่มเติม 87

รายการอ้างอิง 93

ภาคผนวก 97

ภาคผนวก ก 98

ภาคผนวก ข 121

ประวัติผู้เขียนวิทยานิพนธ์ 150

สารบัญตาราง

ตาราง หน้า

3-1 ตัวอย่างข้อมูลที่เก็บมาจากพจนานุกรมศัพท์สัทศาสตร์....46

3-2 ตัวอย่างข้อมูลที่เก็บมาจากศัพท์ภาษาศาสตร์....46

3-3 ตัวอย่างข้อมูลที่จับคู่ข้อมูลแล้ว....49

3-4 ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง....51

3-5 ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง....51

3-6 ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย 3 ตำแหน่งและขวา 3 ตำแหน่ง....52

3-7 ตัวอย่างข้อมูลในตารางคาดเดาที่ไม่อาศัยบริบท....53

3-8 ตัวอย่างข้อมูลในตารางคาดเดาที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง....53

3-9 ตัวอย่างข้อมูลในตารางคาดเดาที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง....53

4-1 ตัวอย่างตารางเทียบเสียงสระภาษาอังกฤษของราชบัณฑิตยสถาน....61

4-2 เปรียบเทียบความสอดคล้องการถอดสระ....64

4-3 ตัวอย่างการถอดสระที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน....65

4-4 เปรียบเทียบความสอดคล้องการถอดพยัญชนะต้น....67

4-5 ตัวอย่างการถอดพยัญชนะต้นที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน....68

4-6 เปรียบเทียบความสอดคล้องการถอดพยัญชนะท้าย....70

4-7 ตัวอย่างการถอดพยัญชนะท้ายที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน....71

4-8 ตัวอย่างคำที่ถอดอักษรไม่สอดคล้องกับกฎของราชบัณฑิตยสถาน....72

4-9 ตัวอย่างการถอดอักษรที่ไม่สอดคล้องเนื่องจากการพบการถอดอักษรที่นอกเหนือไปจากกฎการถอดอักษรของราชบัณฑิตยสถาน....74

4-10 ประสิทธิภาพในการถอดอักษรของโปรแกรมการถอดอักษรในระดับคำ....75

4-11 ประสิทธิภาพในการถอดอักษรของโปรแกรมการถอดอักษรในระดับอักษร....76

4-12 ตัวอย่างคำที่ถอดอักษรถูกต้อง....78

4-13 ตัวอย่างคำที่ถอดอักษรผิดพลาด....78

4-14 ตัวอย่างการถอดอักษรในคลังข้อมูลที่ไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน....79

4-15 ตัวอย่างข้อมูลที่แสดงว่าผู้ถอดอักษรในคลังข้อมูลคำทับศัพท์ของราชบัณฑิตยสถานออกเสียงคำภาษาอังกฤษแตกต่างกัน....81

4-16 ตัวอย่างข้อมูลที่ถอดอักษรผิดเนื่องจากไม่ได้กำหนดขอบเขตพยางค์....82

หน้า

4-17 ตัวอย่างคำที่ถอดอักษรผิดเพราะโปรแกรมเลือกอักษรที่จะนำมาเป็นผลลัพธ์ของการถอดอักษรผิด 83
4-18 ตัวอย่างคำที่ถอดอักษรผิดเพราะการไม่มีรูปแทนเสียงเป็นตัวกลางในการถอดอักษร 84
5-1 กฎการถอดสระเพิ่มเติมจากกฎของราชบัณฑิตยสถาน.....88
5-2 กฎการถอดพยัญชนะเพิ่มเติมจากกฎของราชบัณฑิตยสถาน 89
5-3 ผลการถอดอักษรคำยืมใหม่.....90

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพประกอบ หน้า

3-1 กระบวนการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน.....44

3-2 ลำดับการเทียบตารางกฎและตารางคาดเดา.....55

4-1 ตารางกฎสำหรับคอมพิวเตอร์.....59

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้เป็นยุคแห่งข่าวสารข้อมูล ประเทศไทยได้รับวัฒนธรรมตะวันตกเข้ามามากมาย รวมทั้งมีนวัตกรรมใหม่เพิ่มมากขึ้น อิทธิพลของภาษาต่างประเทศในประเทศไทยจึงมากขึ้นตามไปด้วย โดยเฉพาะภาษาอังกฤษซึ่งเป็นภาษาสากลของคนทั่วโลกนั้นได้เข้ามามีบทบาทในประเทศเรามาก ดังจะเห็นได้ว่ามีคนนำคำทับศัพท์ภาษาอังกฤษมาใช้ในทุกด้าน ทั้งในด้านศัพท์วิชาการและศัพท์เทคนิค ราชบัณฑิตยสถานได้มีการบัญญัติศัพท์วิชาการขึ้นใช้ แม้จะยึดหลักว่าพยายามคิดคำขึ้นใหม่ โดยผสมขึ้นจากคำไทยหรือคำบาลีสันสกฤตที่ใช้อยู่ในภาษาไทยแล้ว ก็ยังมีคำอีกเป็นจำนวนมากที่ไม่สามารถหาคำไทยมาใช้ให้ตรงกับความหมายที่ต้องการได้ จึงต้องใช้วิธีทับศัพท์ นอกจากนี้ การเขียนคำวิสามานยนาม เช่น ชื่อคน ชื่อสถานที่ ก็จำเป็นต้องใช้วิธีการทับศัพท์เช่นเดียวกัน อีกทั้งคำทับศัพท์หลายคำแพร่หลายรวดเร็วมากจนราชบัณฑิตยสถานไม่สามารถบัญญัติศัพท์ได้ทันและบางครั้งศัพท์ที่ราชบัณฑิตยสถานบัญญัติออกมานั้นก็ไม่เป็นที่นิยม ราชบัณฑิตจึงมีการวางหลักเกณฑ์การทับศัพท์ภาษาอังกฤษขึ้นเพื่อให้เกิดมาตรฐานร่วมกันในการทับศัพท์ ดังจะเห็นได้จากหนังสือหลักเกณฑ์การทับศัพท์ภาษาอังกฤษฉบับราชบัณฑิตยสถาน ปี พ.ศ. 2532 แต่ทว่าคนทั่วไปก็มิได้ใช้กฎตามที่ราชบัณฑิตยสถานบัญญัติไว้ในการทับศัพท์ทุกครั้ง ซึ่งอาจเป็นเพราะการเทียบกฎของราชบัณฑิตยสถานที่เขียนไว้ในหนังสือนั้นทำได้ช้าและไม่สะดวก จึงทำให้เกิดปัญหาความหลากหลายของคำทับศัพท์ขึ้น เช่น คำว่า Thomas อาจจะเขียนได้เป็น โธมัส หรือ ทอมัส และคำว่า gas อาจจะเขียนได้เป็น แก๊ส หรือ ก๊าซ

สำหรับวิทยานิพนธ์เล่มนี้จะขอเรียกการทับศัพท์ว่า “การถอดอักษร” ซึ่งแปลมาจากคำภาษาอังกฤษว่า transliteration มีความหมายว่า การแทนตัวอักษรในภาษาหนึ่งด้วยตัวอักษรอีกภาษาหนึ่งซึ่งมีลักษณะทางเสียงคล้ายคลึงกัน และจะพัฒนาโปรแกรมการถอดอักษร เพื่อช่วยทำให้การถอดอักษรตามแนวทางของราชบัณฑิตยสถานทำได้ง่ายและสะดวกรวดเร็วขึ้น ซึ่งหากคนทั่วไปได้ใช้โปรแกรมนี้ในการถอดอักษร ก็จะทำให้การถอดอักษรภาษาอังกฤษเป็นภาษาไทยเป็นไปตามมาตรฐานที่ราชบัณฑิตยสถานวางไว้มากขึ้น ทั้งนี้ผู้วิจัยจะทำการศึกษาการถอดอักษรนี้ตามแนวภาษาศาสตร์คอมพิวเตอร์ ซึ่งก็คือ การศึกษาและเก็บรวบรวมข้อมูลทางภาษา และ

นำมาประยุกต์ใช้กับคอมพิวเตอร์ ในวิทยานิพนธ์นี้ ข้อมูลทางภาษาคือคำภาษาอังกฤษและคำทับศัพท์ภาษาไทยที่ถูกรวบรวมเป็นคลังข้อมูลภาษา* ซึ่งจะนำมาเป็นข้อมูลให้คอมพิวเตอร์สร้างเป็นกฎการถอดอักษรต่อไป

เมื่อพิจารณาจากกฎที่ราชบัณฑิตยสถานบัญญัติไว้นั้นพบว่า หากจะนำกฎที่ราชบัณฑิตยสถานบัญญัติไว้มาประยุกต์ใช้กับคอมพิวเตอร์แล้วจะเกิดปัญหาคือ กฎของราชบัณฑิตยสถานนั้นยังไม่ละเอียดเพียงพอที่จะนำมาเขียนเป็นกฎให้คอมพิวเตอร์เข้าใจได้ เนื่องจากการเขียนโปรแกรมนั้นจำเป็นต้องเขียนบริบทของการเกิดตัวอักษรหนึ่งๆให้ละเอียดชัดเจนว่าตัวอักษรภาษาอังกฤษตัวหนึ่งจะถอดอักษรเป็นตัวอักษรไทยอีกตัวหนึ่งได้เมื่ออยู่ที่ตำแหน่งใด นั่นคือ ข้างหน้า ข้างหลัง หรือระหว่างตัวอักษรใด เมื่อเทียบกับกฎในการเขียนทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานนั้นจะเห็นว่ากฎของราชบัณฑิตยสถานเป็นเพียงกฎคร่าวๆ ซึ่งไม่ได้บอกตำแหน่งที่ชัดเจนของตัวอักษรภาษาอังกฤษไว้ ตัวอย่างที่เห็นได้ชัดเจน คือในกรณีของสระ a ในภาษาอังกฤษ ในตารางเทียบเสียงสระภาษาอังกฤษในหนังสือหลักเกณฑ์การทับศัพท์ภาษาอังกฤษเขียนไว้ดังนี้

สระภาษาอังกฤษ	ใช้สระ	ตัวอย่าง
a	แอ	badminton = แบดมินตัน
	อะ	aluminium = อะลูมิเนียม
	อา	Chicago = ชิคาโก
	เอ	Asia = เอเชีย
	ออ	football = ฟุตบอล

จะเห็นได้ว่าจากตารางข้างต้นนั้นราชบัณฑิตยสถานไม่ได้บอกบริบทที่ชัดเจนของการถอดอักษรสระ a ในภาษาอังกฤษว่า เมื่อใดที่ผู้ใช้ควรจะถอดอักษร a ด้วยสระ แอ หรือเมื่อใดควรถอดอักษร a ด้วยสระ อะ ราชบัณฑิตยสถานเพียงแค่ยกตัวอย่างว่า สระ a ภาษาอังกฤษนั้นสามารถออกเสียงเป็นสระใดในภาษาไทยได้บ้างเท่านั้น ในกรณีนี้แม้แต่คนไทยก็ยังมีความสับสนในการแยกแยะการออกเสียงคำๆหนึ่งในภาษาอังกฤษว่าจะออกเสียงอย่างไรจะถูกต้อง อีกทั้งการออก

* คลังข้อมูลภาษา (corpus) คือ ข้อมูลภาษาเขียนหรือภาษาพูดที่เป็นภาษาที่ใช้จริงซึ่งถูกรวบรวมขึ้นมาในปริมาณที่มากเพียงพอตามข้อกำหนดหรือเงื่อนไขที่กำหนดขึ้น เพื่อนำคลังข้อมูลนั้นมาใช้ประโยชน์ในการศึกษาเรื่องต่างๆที่เกี่ยวข้องกับภาษา (วิโรจน์ อรุณมานะกุล 2545: 1)

เสียงภาษาอังกฤษนั้นก็มีทั้งการออกเสียงแบบสำเนียงอังกฤษและอเมริกัน ทำให้สระ a เพียง 1 ตัว นั้นสามารถออกเสียงได้หลายเสียง ในการที่จะแก้ปัญหานี้เพื่อเขียนกฎให้คอมพิวเตอร์เข้าใจได้นั้น ผู้วิจัยจึงทำการรวบรวมคลังข้อมูลคำทับศัพท์ภาษาอังกฤษจากหนังสือที่ราชบัณฑิตยสถานได้จัดตีพิมพ์ไว้และจะนำคลังข้อมูลที่รวบรวมมาได้นี้มาสร้างเป็นกฎที่ละเอียดเพียงพอที่จะนำไปใช้กับคอมพิวเตอร์ได้ และจะพัฒนาโปรแกรมที่ใช้ในการถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยขึ้นมาโดยยึดหลักเกณฑ์ตามที่ราชบัณฑิตยสถานได้วางเอาไว้เพื่อเป็นประโยชน์กับคนทั่วไปที่ต้องการถอดอักษรให้ได้อย่างถูกต้องตามกฎของราชบัณฑิตยสถาน

1.2 **วัตถุประสงค์ของการวิจัย**

1.2.1 สร้างกฎการถอดอักษรภาษาอังกฤษเป็นภาษาไทยสำหรับใช้กับคอมพิวเตอร์โดยอาศัย ข้อมูลจากคลังข้อมูลคำทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานที่รวบรวมไว้

1.2.2 เปรียบเทียบกฎที่ราชบัณฑิตยสถานได้วางไว้กับกฎที่ได้จาการสร้างด้วยคอมพิวเตอร์ว่ามีความสอดคล้องกันเพียงใด

1.2.3 พัฒนาโปรแกรมการถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยขึ้นมาโดยยึดหลักเกณฑ์ตามที่ราชบัณฑิตยสถานให้ไว้ โดยคาดหวังความถูกต้อง ไม่ต่ำกว่าร้อยละ 90

1.3 **สมมติฐานในการวิจัย**

1.3.1 กฎการถอดอักษรภาษาอังกฤษเป็นภาษาไทยสามารถสร้างขึ้นมาได้โดยอาศัยคลังข้อมูลคำทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานที่รวบรวมไว้ได้

1.3.2 กฎการถอดอักษรที่ราชบัณฑิตยสถานวางไว้มีความสอดคล้องกับกฎการถอดอักษรที่สร้างขึ้นมาสำหรับใช้กับคอมพิวเตอร์

1.3.3 สามารถพัฒนาโปรแกรมการถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยขึ้นมาโดยยึดหลักเกณฑ์ตามที่ราชบัณฑิตยสถานให้ไว้ โดยมีความถูกต้องไม่ต่ำกว่า ร้อยละ 90

1.4 กรอบทฤษฎี

ประยุกต์ใช้แบบจำลอง (Model) แบบเปรียบเทียบตารางกฎ (Table Lookup) ตามแนวของ Van den Bosch and Daelemans (1993) ที่ใช้วิธีนำข้อมูลที่เก็บรวบรวมมาสร้างเป็นตารางกฎ เพื่อให้คอมพิวเตอร์นำไปประมวลผลในการถอดอักษร

1.5 เครื่องมือที่ใช้ในงานวิจัย

1.5.1 เครื่องคอมพิวเตอร์ส่วนบุคคล

1.5.2 โปรแกรมภาษา Perl ของบริษัท Active Perl

1.6 ขอบเขตของการวิจัย

ศึกษาเฉพาะคำทับศัพท์ภาษาอังกฤษที่ถอดอักษรเป็นไทยที่ตีพิมพ์หรือเผยแพร่โดยราชบัณฑิตยสถานตั้งแต่ปี พ.ศ. 2508 จนถึงปัจจุบัน จำนวนไม่ต่ำกว่า 10,000 คำ

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 ได้โปรแกรมการถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมาตามเกณฑ์ของราชบัณฑิตยสถาน

1.7.2 ช่วยให้การถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยเป็นไปอย่างมีมาตรฐานเดียวกันตามเกณฑ์ราชบัณฑิตยสถานมากขึ้น

1.8 วิธีดำเนินการวิจัย

1.8.1 ทำการเก็บรวบรวมคลังข้อมูลคำทับศัพท์ภาษาอังกฤษที่ทับศัพท์ตามเกณฑ์ราชบัณฑิตยสถานจากหนังสือและเว็บไซต์ของราชบัณฑิตยสถานต่างๆที่ราชบัณฑิตยสถานจัด

ตีพิมพ์ โดยจัดเก็บเป็นคู่คำภาษาอังกฤษและคำไทย จับเก็บในโปรแกรม Microsoft Excel เช่น Abbasid /อับบาซิด จนได้คลังข้อมูลที่มีขนาดประมาณ10,000 คำ

1.8.2 เขียนโปรแกรมจับคู่ตัวอักษรภาษาอังกฤษและภาษาไทยเพื่อให้ตัวอักษรไทยและอังกฤษในตำแหน่งเดียวกันจับคู่กัน เช่น a/b/b/a/s/i/d/ อั/บ/บ/า/ซิ//ด/

1.8.3 เขียนโปรแกรมให้จับข้อมูลเป็นคู่ๆที่ได้มาบรรจุลงในตารางที่เรียงลำดับตามบริบทการเกิดของตัวอักษร โดยแบ่งออกเป็นตาราง 2 ประเภท คือ ตารางกฎและตารางคาดเดา โดยแยกเป็นอย่างละ 3 ตารางย่อย ดังนี้

1) ตารางกฎที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง

2) ตารางกฎที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง

3) ตารางกฎที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 3 ตำแหน่งและขวา 3 ตำแหน่ง

สำหรับอักษรที่ไม่สามารถสรุปเป็นกฎในตารางทั้งสามได้ หรืออักษรที่สามารถถอดอักษรได้หลายแบบในบริบทเดียวกัน ก็จะเลือกแบบที่ปรากฏมากที่สุดเก็บไว้ในตารางคาดเดา 3 ตารางที่ช่วยในการคาดเดาตัวอักษรที่จะถอดออกมา ดังนี้

1) ตารางคาดเดาโดยไม่ต้องใช้บริบท

2) ตารางคาดเดาที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง

3) ตารางคาดเดาที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง

ตารางคาดเดาทั้งสามนี้ จะใช้ในกรณีที่อักษรที่รับเข้ามาไม่ตรงกับกฎใดๆในตารางกฎทั้งสาม

1.8.4 นำกฎที่ได้จากตารางนี้มาเปรียบเทียบดูความสอดคล้องกับกฎที่ราชบัณฑิตยสถานบัญญัติไว้ว่าสอดคล้องกันมากเพียงใด

1.8.5 เขียนโปรแกรมที่ใช้ในการเทียบคำภาษาอังกฤษที่รับเข้ามากับข้อมูลที่มีอยู่ในตาราง ซึ่งเมื่อพบข้อมูลซึ่งมีบริบทตรงกับในตารางก็จะให้คำภาษาไทยที่ตรงตามกฎออกมา โดยคาดหวังความถูกต้องไม่น้อยกว่าร้อยละ 90

1.9 **โครงสร้างของวิทยานิพนธ์**

วิทยานิพนธ์เล่มนี้ประกอบด้วยเนื้อหา 5 บท ในบทที่ 1 ผู้วิจัยจะนำเสนอเรื่องความเป็นมาของปัญหาที่นำมาซึ่งการทำวิทยานิพนธ์ฉบับนี้ กล่าวถึงวัตถุประสงค์ ประโยชน์ และวิธีดำเนินการวิจัยอย่างย่อ ต่อมาในบทที่ 2 จะนำเสนอเกี่ยวกับเอกสารและงานวิจัยที่เกี่ยวข้องกับแนวทางการถอดอักษรทั้งในด้านภาษาศาสตร์และวิธีการถอดอักษรโดยใช้คอมพิวเตอร์ ส่วนบทที่ 3 จะ

นำเสนอเกี่ยวกับวิธีการดำเนินการวิจัยและการวิเคราะห์ข้อมูลโดยละเอียด สำหรับบทที่ 4 จะนำเสนอเกี่ยวกับผลการถอดอักษรโดยใช้คอมพิวเตอร์ สุดท้ายในบทที่ 5 จะนำเสนอเกี่ยวกับการสรุปผลการวิจัยการอภิปรายผลและข้อเสนอแนะ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

บทนี้จะนำเสนอเกี่ยวกับประวัติความเป็นมาของการถอดอักษรในภาษาไทย คำจำกัดความของการถอดอักษร (Transliteration) ที่มีหลายคนสับสนกับคำว่าการถ่ายเสียง (Transcription) ปัญหาในการถอดอักษร และอธิบายเกี่ยวกับเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน นอกจากนี้จะนำเสนอเกี่ยวกับแนวทางการถอดอักษรในภาษาต่างๆโดยใช้คอมพิวเตอร์ซึ่งได้มีผู้ทำมาแล้ว

2.1 ประวัติความเป็นมาของการถอดอักษรในภาษาไทย

กาญจนา นาคสกุล (ม.ป.ป.: 1) กล่าวว่า ชนชาวอังกฤษได้เริ่มเข้ามาติดต่อกับประเทศไทยตั้งแต่สมัยกรุงศรีอยุธยาโดยส่งเรือเข้ามายังปัตตานี และส่งทูตเข้ามาถึงกรุงศรีอยุธยา แต่ภาษาไทยเริ่มรับคำจากภาษาอังกฤษเข้ามาใช้ตั้งแต่สมัยพระบาทสมเด็จพระนั่งเกล้าเจ้าอยู่หัว เมื่อเริ่มมีการติดต่อทางการทูตและการค้ากับอังกฤษ คำภาษาอังกฤษที่เข้ามาในช่วงแรกเป็นคำที่เรียกตำแหน่งข้าราชการต่างๆ ต่อมาในสมัย รัชกาลที่ 4 โปรดเกล้าฯให้จ้างฝรั่งมาสอนภาษาอังกฤษให้แก่พระราชโอรสพระราชธิดา

คำภาษาอังกฤษที่รับเข้ามาในสมัยแรกๆมักมีการแปลงเสียงในลักษณะลากเข้าเสียงคำในภาษาไทย เช่น กัดฟันมัน สเตแท่น แร้งกิน ฝาศุภเรส หันแตร บารนี ฯลฯ ซึ่งในปัจจุบันคำเหล่านี้ไม่มีใช้ในภาษาไทยแล้ว แต่คำที่รับมาในสมัยต่อมาและยังคงใช้อยู่ในภาษาไทยยังมีอยู่ เช่น กาว กิ๊บ เก๋ โกดัง คัดชู จิ๊กโก๋ ซุป ฯลฯ

กาญจนา นาคสกุล (ม.ป.ป.: 1) กล่าวถึงการเขียนคำทับศัพท์ภาษาอังกฤษเพิ่มเติมว่าภาษาไทยยืมคำจากภาษาอื่นมาใช้หลายภาษาและมีจำนวนมากมาย เช่น ยืมจากภาษาบาลี ภาษาสันสกฤตภาษาเขมร ภาษาจีน ภาษาชวา ภาษามลายู ภาษาญี่ปุ่น ภาษาเปอร์เซีย ภาษามอญ ภาษาเวียดนาม ซึ่งล้วนเรียกว่า “คำยืม” มีเฉพาะคำที่ยืมมาจากภาษาอังกฤษเท่านั้นที่ปัจจุบันมักเรียกว่า “คำทับศัพท์” อาจเป็นเพราะ ปัจจุบันมีการเรียนภาษาอังกฤษ เมื่อใช้คำภาษาอังกฤษก็มิได้รู้สึกว่ายืมคำมาใช้ในภาษาไทย แต่รู้สึกคล้ายว่าคำภาษาอังกฤษนั้นเป็นคำที่สื่อสารกันในภาษาไทย เพียงแต่เมื่อเป็นคำภาษาอังกฤษ ก็หาวิธีเขียนออกมาเป็นตัวอักษรไทย ผู้

ยืมคำภาษาอังกฤษบางคนเขียนคำภาษาอังกฤษด้วยตัวอักษรโรมัน แต่เพื่อให้ผู้ที่ไม่ชินกับอักษรโรมันอ่านได้ หรือเพื่อมิให้ข้อเขียนมีตัวอักษรหลายระบบปนกัน จึงได้ถ่ายคำภาษาอังกฤษในข้อเขียนนั้นออกเป็นอักษรไทย ด้วยวิธีการเขียนตามเสียงบ้าง เขียนถ่ายอักษรทีละตัวแบบที่เรียกว่า Transliteration บ้าง และการเขียนคำภาษาอังกฤษก็มักจะสะท้อนตัวอักษรที่เขียนคำนั้นคำยืมจากภาษาอังกฤษจึงมักเรียกว่า “คำทับศัพท์”

บุญเสริม ฤทธาธิรมย์ (2522: 2) กล่าวว่า ภาษาอังกฤษเริ่มมีอิทธิพลต่อภาษาไทยนับแต่สมัยรัชกาลที่ 3 เป็นต้นมา โดยเฉพาะรัชกาลที่ 4 ทรงสนพระทัยและศึกษาภาษาอังกฤษอย่างกว้างขวาง โปรดเกล้าฯให้จ้างฝรั่งมาสอนภาษาอังกฤษให้แก่พระราชโอรสพระราชธิดา และเชื้อพระวงศ์ทั้งหลาย กษัตริย์องค์ต่อมาก็ทรงเห็นความสำคัญของภาษาอังกฤษ มีการส่งพระโอรสและนักเรียนทุนหลวงไปศึกษาต่อต่างประเทศ มีการยอมรับความช่วยเหลือจากต่างประเทศ เช่น ยอมรับความช่วยเหลือจากผู้เชี่ยวชาญ มิชชันนารี นักวิชาการ จึงเกิดความจำเป็นในการเรียน การสอนภาษาอังกฤษ ในระดับโรงเรียน ภาษาอังกฤษจึงกลายเป็นวิชาหนึ่งของหลักสูตรในโรงเรียนตั้งแต่ชั้นต้นไปจนถึงระดับอุดมศึกษามาจนทุกวันนี้

เมื่อมีการติดต่อกับชนชาติที่ใช้ภาษาอังกฤษ และคนไทยรู้ภาษาอังกฤษก็ก่อให้เกิดการเปลี่ยนแปลงในภาษาไทย โดยเฉพาะอย่างยิ่งมีคำภาษาอังกฤษหลายคำที่เข้ามาปะปนกับคำไทย ซึ่งกลายเป็นภาษาที่เกิดใหม่ขึ้น คนไทยนำมาใช้โดยไม่มีการแปลหรือบัญญัติเป็นคำไทย เปรียบกับคนก็คือพวกต่างด้าวที่เข้ามาอยู่ในเมืองไทย แล้วโอนสัญชาติเป็นไทยฉันนั้น คำต่างชาติที่นำมาใช้นี้เรียกว่าคำที่ยืมมาใช้ (borrowed term หรือ loan-word) เป็นลักษณะการใช้แบบทับศัพท์ (transliteration) การยืมคำมาใช้แบบทับศัพท์ก็คือการถ่ายเสียงคำภาษาอังกฤษเป็นภาษาไทย โดยการเลียนเสียงอักษรของแต่ละภาษาให้ตรงกัน หรือใกล้เคียงกัน หรือออาจจะเปลี่ยนเสียงเสียใหม่ตามความสะดวกในการออกเสียงและการสะกดในภาษาไทย

2.1.1 **สาเหตุของการใช้คำทับศัพท์**

บุญเสริม ฤทธาธิรมย์ (2522: 3) กล่าวว่าการยืมคำภาษาอังกฤษมาใช้แบบทับศัพท์น่าจะเป็นเพราะสาเหตุดังต่อไปนี้

2.1.1.1 **ไม่มีคำนั้นในภาษาไทย** เช่น ชื่อเฉพาะต่างๆ หน่วยการชั่งตวงวัด ชื่อพืช สัตว์ ต้นไม้ ดนตรี อาหาร เครื่องแต่งตัว ฯลฯ เช่นคำว่า ละติน โอลิมปิค ทาร์ซาน ฟุต ไมล์ ลิตร กิโลกรัม อะลูมิเนียม กีวี เบสบอล สกี ฯลฯ

2.1.1.2 **เป็นศัพท์ที่ใช้กันในเฉพาะสาขาวิชา หรือศัพท์เทคนิค** (technical terms) ก่อนนั้นอาจจะเป็นศัพท์ที่ใช้กันในกลุ่มคนบางอาชีพ เป็นที่รู้กันหรือเข้าใจกันในบุคคลที่มีอาชีพอย่างเดียวกัน ตัวอย่างเช่น ในวิชาอาหารโภชนาการ มีคำว่า แคลอรี คาร์โบไฮเดรต โปรตีน วิตามิน ฯลฯ สาขาวิทยาศาสตร์ เช่น นิวเคลียร์ กาแลกซี พลาสติก โครโมโซม โปรตอน นิวตรอน ฯลฯ ทางการแพทย์ เช่นคำว่า โปลิโอ แอสไพริน เฮโรอีน ฯลฯ แต่ต่อมาคำเหล่านี้กลายเป็นที่นิยมใช้กันอย่างกว้างขวางในสาขาวิชาอื่นๆ และของคนทั่วไป ถ้าจะแปลเป็นไทย อาจจะสร้างความหมายที่ต่างออกไป

2.1.1.3**อาจเป็นเพราะการใช้คำทับศัพท์ใช้อธิบายความหมายในภาษาไทยได้กว้างขวางกว่า** เท่ากับเป็นการประหยัดคำอธิบายในภาษาไทยด้วยในตัว เช่นคำว่า คอร์รัปชันนั้น เพียงคำเดียว สามพยางค์ ดูจะง่าย รัดกุม ประหยัดคำมากกว่าที่จะไปอธิบายว่า “การกระทำความชั่วในลักษณะต่างๆ การฉ้อราษฎร์บังหลวง”

2.1.1.4 **ไม่มีการบัญญัติศัพท์หรือแปลเป็นภาษาไทย** หรือแม้จะถ่ายความหมายเป็นไทยแล้ว บางคำก็เป็นที่เข้าใจยาก จะต้องอธิบายศัพท์บัญญัติหรือคำแปลนั้นๆ เป็นภาษาไทยอีกต่อหนึ่ง การใช้แบบทับศัพท์จึงสะดวกกว่า เช่น

โควตา (quota) - ส่วนกฎเกณฑ์ (กรมหมื่นนราธิปพงศ์ประพันธ์)

ซุปเปอร์แมน (superman) - อภิมนุษย์ (ราชบัณฑิตยสถาน)

โรแมนติก (romantic) - จินตนิยม (ราชบัณฑิตยสถาน)

ฯลฯ

2.1.1.5 **การใช้ภาษาต่างประเทศหรือการทับศัพท์ แสดงถึงเกียรติภูมิอย่างหนึ่งของผู้ใช้** ทำให้เกิดความรู้สึกว่าผู้ใช้มีความแตกต่างจากคนอื่น เป็นคนมีความรู้ หรือเพื่อไม่ให้คนกลุ่มอื่นสื่อความหมายได้ บางทีก็เป็นลักษณะนิสัยติดตัวของบางคนที่ชอบใช้ภาษาอังกฤษปะปนกับคำไทย เช่น คนที่มีความรู้ภาษาอังกฤษ คนที่สำเร็จจากต่างประเทศ นักวิชาการ ฯลฯ เช่น ประโยคต่อไปนี้

วันนี้เขามีนัดกับคนรัก ใช้ว่า วันนี้เขามีเดทกับแฟน

เขาไม่สนใจเรื่องที่ถูกกล่าวหา ใช้ว่า เขาไม่แคร์เรื่องที่ถูกกล่าวหา

เขาต่อต้านเรื่องนี้อย่างมาก ใช้ว่า เขาแอนตี้เรื่องนี้อย่างมาก

ฯลฯ

2.1.2 **ลักษณะการใช้การทับศัพท์**

ดังที่กล่าวมาข้างต้นแล้วว่า การทับศัพท์เป็นลักษณะการยืมคำภาษาต่างประเทศมาใช้ การทับศัพท์อาจทำได้หลายอย่าง สุทธิวงศ์ พงศ์ไพบูลย์ (2516: 15) ได้แยกประเภทของลักษณะการทับศัพท์ไว้ 3 พวก ดังนี้

2.1.2.1 **ทับศัพท์โดยตรง** (transliteration) เป็นการทับศัพท์ แบบถ่ายเสียงอักษรมาตัวต่อตัว พยายามเลียนเสียงอักษรของแต่ละภาษาให้ตรงกันหรือใกล้เคียงกันมากที่สุด เช่น คำว่า ฟุตบอล รีม คิว วิว ทีนเอจ ฟุต ไมล์ ดีกรี ตัน ฯลฯ

2.1.2.2 **ทับศัพท์โดยตัดเสียงหรือเปลี่ยนเสียง** เพื่อความสะดวกในการออกเสียง และการสะกดในภาษาไทย ตัวอย่างเช่นคำว่า เซ็น (sign) สาลี่ (trolley) จิ๊กโก๋ (gigolo) ฯลฯ

2.1.2.3 **โดยการถอดเป็นคำใหม่ แปลเอาความ** (loan translation) ได้แก่การแปลเป็นคำใหม่ โดยพยายามรักษาเค้าความหมายตามรากศัพท์เดิมไว้ เช่น

telephone	เป็น	โทรศัพท์
seminar	เป็น	สัมมนา
television	เป็น	โทรทัศน์

สำหรับในวิทยานิพนธ์เล่มนี้ จะศึกษาเฉพาะคำทับศัพท์โดยตรง (Transliteration) และจะใช้คำเรียกการทับศัพท์โดยตรงว่า “การถอดอักษร” ซึ่งมักจะมีผู้สับสนกับ “การถ่ายเสียง” จึงจะกล่าวถึงคำนิยามของทั้งสองคำนี้ในหัวข้อถัดไป

2.2 คำจำกัดความ ความแตกต่างและลักษณะเฉพาะของการถอดอักษร (Transliteration) และ การถ่ายเสียง (Transcription)

เมื่อกล่าวถึงการถอดอักษร (Transliteration) แล้ว มักจะมีคนสับสนกับการถ่ายเสียง (Transcription) คำสองคำนี้มีความแตกต่างกันดังนี้

2.2.1 คำจำกัดความของการถอดอักษร

บลูมฟิลด์ (Bloomfield 1974: 90) กล่าวว่า ระบบการถอดอักษรประกอบด้วยการกำหนดให้อักษรลาติน (หรือกลุ่มอักษรอื่น) ตรงกับตัวเขียนในภาษาเดิมทำให้สามารถเขียนภาษาเดิมนั้นได้ด้วยอักษรลาติน นั่นคือ การถอดอักษรของภาษาหนึ่งให้เป็นอักษรในอีกภาษาหนึ่ง โดยกำหนดให้กลุ่มอักษร หรือกลุ่มสัญลักษณ์ ตรงกับอักษรในภาษาเดิมและทำให้สามารถเขียนคำในภาษานั้นได้ด้วยอักษรของอีกภาษาหนึ่ง

แคทฟอร์ด (Catford, 1965: 50-66) กล่าวว่าระบบการถอดอักษร คือ การถอดอักษรจากภาษาหนึ่งเป็นอักษรของอีกภาษาหนึ่งโดยมีกฎในการถอดอักษร เขาได้กล่าวถึงหลักใหญ่ๆในการถอดอักษรซึ่งมี 3 ขั้นตอน คือ

1) ตัวอักษรภาษาที่ 1 ถูกถ่ายเสียงให้เป็นหน่วยเสียงในภาษานั้น เช่น
หน่วยอักษรภาษา ก. “b” ถ่ายเป็นหน่วยเสียงภาษา ก. /b/

2) หน่วยเสียงในภาษาที่ 1 เทียบให้เป็นหน่วยเสียงในภาษาที่ 2 เช่น
หน่วยเสียงภาษา ก. /b/ เทียบเป็นหน่วยเสียงภาษา ข. /b/

3) ถอดหน่วยเสียงในภาษาที่ 2 เป็นอักษรในภาษาที่ 2 เช่น
หน่วยเสียงภาษา ข. /b/ ถอดหน่วยเสียงเป็นอักษรในภาษา ข. “บ”

สามารถแสดงด้วยภาพได้ดังนี้

L1 L2

“b” -> /b/ <-> /b/ -> “บ”

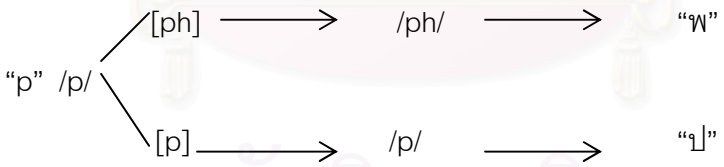
กล่าวโดยสรุปคือ ระบบการถอดอักษร (Transliteration) คือ ระบบการถอดอักษรภาษาหนึ่งด้วยอักษรของอีกภาษาหนึ่งแบบอักษรต่ออักษรโดยพยายามให้ได้เสียงใกล้เคียงกันที่สุด โดยมีแนวในการถอดอักษร อยู่ 2 แนวได้แก่

1) การถอดอักษรโดยยึดอักษรเป็นสำคัญ ได้แก่ การถอดอักษรแบบอักษรต่ออักษร เช่น

อักษรภาษาอังกฤษ	**หน่วยเสียง ภาษาอังกฤษ**	**หน่วยเสียง ภาษาไทย**	**อักษรภาษาไทย**
“b”	/b/	/b/	“บ”
“p”	/p/	/ph/	“พ”

จากตารางนี้อักษร “b” ซึ่งมีหน่วยเสียง (phoneme) /b/ ในภาษาอังกฤษจะถอดอักษรได้เป็นอักษร “บ”ซึ่งมีหน่วยเสียง /b/ ในภาษาไทย ส่วนอักษร “p” ซึ่งมีหน่วยเสียง /p/ ในภาษาอังกฤษจะถอดอักษรได้เป็นอักษร “พ” ซึ่งมีหน่วยเสียง “ph” ในภาษาไทย อักษร “p” นั้นจะถอดอักษรเป็น “พ” ไม่ว่าเสียงจริงจะเป็นอย่างไร การถอดอักษรแนวนี้จึงยึดรูปอักษรเป็นสำคัญ

2) การถอดอักษรโดยยึดเสียงเป็นสำคัญ ได้แก่ การถอดอักษรโดยพยายามให้การถอดอักษรได้เสียงใกล้เคียงที่สุด เช่น



จากข้อมูลด้านบนอักษร p ซึ่งมีหน่วยเสียง /p/ ในภาษาอังกฤษนั้นมี 2 เสียงย่อย (allophone) คือ [ph] และ [p] สำหรับเสียงย่อย [ph] เมื่อเทียบเสียงแล้วจะถอดอักษรได้เป็นอักษร “พ” ซึ่งมีหน่วยเสียง /ph/ ในภาษาไทย และสำหรับเสียงย่อย [p] เมื่อเทียบเสียงแล้วจะถอดอักษรได้เป็นอักษร “ป” ซึ่งมีหน่วยเสียง /p/ ในภาษาไทย

หากจำแนกการถอดอักษรตามหลักภาษาศาสตร์แล้วอาจจำแนกเป็น 2 ประเภทใหญ่ๆคือ

1) ประเภทยึดเรื่องเสียงเป็นสำคัญโดยไม่คำนึงถึงตัวอักษร ลักษณะการถอดอักษรเช่นนี้เป็นการถอดอักษรแบบถ่ายเสียง เช่น

“c,s,sc,ps” /s/ = /s/ “ส”

นั่นคืออักษร “c,s,sc,ps” ที่มีหน่วยเสียง /s/ ในภาษาอังกฤษจะถอดอักษรเป็นอักษร “ส” ที่มีหน่วยเสียง /s/ ในภาษาไทย ตัวอย่างเช่น การถอดอักษรภาษาไทยเป็นโรมันแบบถ่ายเสียงของราชบัณฑิตยสถาน มีการยึดหลักว่าในการถอดอักษรภาษาไทยเป็นโรมันนั้น จะต้องพิจารณาคุณสมบัติทางเสียงของอักษรไทยและอักษรโรมันที่คล้ายคลึงกันเป็นสำคัญ

2) ประเภทยึดเรื่องตัวอักษรเป็นสำคัญ กล่าวคือจะถอดอักษรแบบอักษรต่ออักษร เช่น

“c” /s/ = /s/ “ศ”
“s” /s/ = /s/ “ส”
“sc” /s/ = /s/ “ส์ซ”

นั่นคืออักษร “c” ที่มีหน่วยเสียง /s/ ในภาษาอังกฤษจะถอดอักษรเป็นอักษร “ศ” ที่มีหน่วยเสียง /s/ ในภาษาไทย อักษร “s”ที่มีหน่วยเสียง /s/ ในภาษาอังกฤษจะถอดอักษรเป็นอักษร “ส” ที่มีหน่วยเสียง /s/ ในภาษาไทย ส่วนอักษร “sc” ที่มีหน่วยเสียง /s/ ในภาษาอังกฤษจะถอดอักษรเป็นอักษร “ส์ซ” ที่มีหน่วยเสียง /s/ ในภาษาไทย เช่น วารสารศิริราชที่ใช้วิธีการถอดอักษรแบบอักษรต่ออักษร โดยที่อักษรภาษาอังกฤษแต่ละตัวจะมีถอดอักษรเป็นภาษาไทยแบบหนึ่งต่อหนึ่ง

ระบบการถอดอักษรที่กำหนด อาจไม่จำเป็นต้องเป็นแบบใดแบบหนึ่งเพียงอย่างเดียว แต่อาจเป็นการผสมของทั้ง 2 ประเภทก็ได้ คือ ยึดทั้งเสียงและตัวอักษรเป็นสำคัญ เช่นการทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถาน ที่กำหนดกฎการถอดอักษรโดยพิจารณาทั้งคุณสมบัติทางเสียงที่คล้ายคลึงกัน เช่น กำหนดให้อักษร “i” ที่มีหน่วยเสียง /I/ ถอดอักษรเป็น “สระอิ” ที่มีหน่วยเสียง /I/ในภาษาไทย อีกทั้งยังยึดตัวอักษรเป็นสำคัญด้วย เช่น กำหนดให้อักษร ‘s’ ที่ตามด้วยสระให้ถอดอักษรเป็น ‘ซ’ แต่ อักษร ‘s’ ที่ตามด้วยพยัญชนะให้ถอดอักษรเป็น ‘ส’ เป็นต้น

2.2.2 คำจำกัดความของการถ่ายเสียง

บลูมฟิลด์ (Bloomfield 1964: 85) กล่าวว่าการถ่ายเสียงเป็นการจดบันทึกข้อมูลทางด้านเสียงของนักภาษาศาสตร์ โดยใช้ระบบสัญลักษณ์ 1 ตัวแทนเสียง 1 เสียง หรือหน่วยเสียง 1 หน่วยเสียงในภาษานั้น สัญลักษณ์แทนเสียงหรือหน่วยเสียงนี้ เรียกว่า “สัทอักษร”

ธีระพันธ์ เหลืองทองคำ (2526: 16-17) กล่าวว่า การถ่ายเสียงหรือการบันทึกข้อมูลทางภาษาโดยใช้สัทอักษร (Phonetic transcription) แตกต่างกับการเขียนภาษาไทยโดยใช้ระบบการเขียน (Writing system) ที่กำหนดไว้ตายตัวแล้ว เวลาบันทึกข้อมูลทางภาษาโดยเฉพาะอย่างยิ่งข้อมูลเกี่ยวกับเสียง ผู้บันทึกจะต้องมีความรู้พื้นฐานทางด้านสัทศาสตร์เพียงพอว่า เสียงที่ตนได้ยินนั้นเป็นเสียงอะไร จะได้เลือกสัทสัญลักษณ์ หรือสัญลักษณ์แทนเสียงได้อย่างถูกต้อง ไม่ว่าจะใช้สัทอักษรสากล หรือสัทอักษรไทยก็ตาม และการถ่ายเสียงโดยใช้สัทอักษรแบ่งออกเป็นแบบใหญ่ๆ 2 แบบ คือ การถ่ายเสียงอย่างละเอียด (Narrow phonetic transcription) และการถ่ายเสียงอย่างมีระบบ (Broad phonetic transcription)

จากคำจำกัดความข้างต้นสรุปได้ว่าการถ่ายเสียงมีลักษณะดังนี้

1) ระบบการถ่ายเสียงเป็นระบบการเขียนที่ถ่ายเสียงในภาษาโดยใช้สัญลักษณ์ 1 ตัว แทนเสียง 1 เสียง หรือ หน่วยเสียง 1 หน่วยเสียง

2) นักภาษาศาสตร์ใช้ระบบการถ่ายเสียงในการจดบันทึกเสียงในภาษา เพื่อเป็นข้อมูลในการศึกษา และวิเคราะห์ภาษานั้นๆ

3) ภาษาทุกภาษานักภาษาศาสตร์สามารถถ่ายเสียง (Transcribe) ได้ แม้ว่าภาษานั้นจะไม่มีระบบการเขียน เพราะระบบการถ่ายเสียงเป็นการกำหนดสัญลักษณ์หรือรูปเขียนใดๆขึ้นแทนเสียง

กล่าวโดยสรุปเกี่ยวกับความแตกต่างระหว่างการถอดอักษร(Transliteration) และ การถ่ายเสียง (Transcription) คือ “การถอดอักษร” เป็นการแทนตัวอักษรในภาษาหนึ่งด้วยตัวอักษรในอีกภาษาหนึ่งที่มีลักษณะทางเสียงคล้ายคลึงกัน แต่ “การถ่ายเสียง”นั้นคือการแทนตัวอักษรในภาษาหนึ่งด้วยรูปแทนเสียงของภาษานั้นซึ่งสามารถแทนได้ด้วยสัทอักษร

ในหัวข้อถัดไปจะขอกล่าวถึงปัญหาทางภาษาศาสตร์และปัญหาอื่นๆในการการถอดอักษร ดังนี้

2.3 ปัญหาทางภาษาศาสตร์และปัญหาอื่น ๆในการถอดอักษร

สำหรับปัญหาทางภาษาศาสตร์และปัญหาอื่นๆในการถอดอักษรนั้น อุไรรัตน์ บุญภานนท์ (2529) ได้สรุปไว้ว่าปัญหาในการถอดอักษรนั้นมีดังนี้

2.3.1 ปัญหาทางภาษาศาสตร์ในการถอดอักษร

2.3.1.1 ปัญหาอันเนื่องมาจากระบบอักขระของภาษาต้นแบบ (Source Language) ที่ความสัมพันธ์ของอักษรและหน่วยเสียงในภาษานั้นๆไม่ได้เป็นแบบ 1 อักษรแทน 1 หน่วยเสียงหรืออักษร 1 อักษรแทนได้หลายเสียง ความสัมพันธ์ของอักษรและหน่วยเสียงนั้นยังมีแบบอักษรหลายตัวแทนหน่วยเสียงเดียว เช่น การถอดอักษรอังกฤษเป็นไทย ซึ่งมีภาษาอังกฤษเป็นภาษาต้นแบบ ตัวอย่างอักษรหลายตัวแทนหน่วยเสียงเดียวในภาษาอังกฤษเช่น “g,gh,gu” แทน /g/, “n,tn,gn,pn” แทน /n/, “r,rh,wr” แทน /r/ เป็นต้น

2.3.1.2 ปัญหาอันเนื่องมาจากระบบอักขระของภาษาเป้าหมาย (Target Language) ความสัมพันธ์ของอักษรและหน่วยเสียงก็อาจมีหลายแบบเหมือนในภาษาต้นแบบ เช่น การถอดอักษรอังกฤษเป็นไทยซึ่งมีภาษาอังกฤษเป็นภาษาต้นแบบและมีภาษาไทยเป็นภาษาเป้าหมาย ระบบอักษรไทยก็มีแบบอักษรหลายตัวแทนหน่วยเสียงเดียว เช่น “ร,ฤ,หร” แทน /r/ “ผ,พ,ภ” แทน /ph/ “ฉ,ช,ฌ” แทน /ch/ เป็นต้น

2.3.1.3 ปัญหาในการเทียบหน่วยเสียงที่ใกล้เคียงกันของทั้งสองระบบ ปัญหานี้จะเกิดถ้าเสียงย่อยในภาษาต้นแบบมีสัทลักษณะที่สามารถแยกเป็นหน่วยเสียงของภาษาเป้าหมายที่แตกต่างกันได้ ตัวอย่างเรื่องปัญหาของเสียงย่อย เช่น หน่วยเสียง /p, t, k/ ในภาษาอังกฤษมีเสียงย่อยที่มีลักษณะพ่นลมในตำแหน่งพยัญชนะต้น [ph, th, kh] ซึ่งเทียบได้ใกล้เคียงกับหน่วยเสียง /ph, th, kh/ ในภาษาไทย และหน่วยเสียง /p, t, k/ ในภาษาอังกฤษก็มีเสียงย่อยเป็น [p, t, k] ในตำแหน่งหลัง “s”- ซึ่งเทียบได้กับหน่วยเสียง /p, t, k/ ในภาษาไทย ดังนั้นในการถอดอักษรอังกฤษเป็นไทยจำต้องคำนึงถึงเสียงย่อยของบางหน่วยเสียงในภาษาอังกฤษที่มีสัทลักษณะที่ใช้แยกหน่วยเสียงในภาษาไทยด้วย

2.3.1.4 ปัญหาที่มีหน่วยเสียงในภาษาต้นแบบแต่ไม่มีหน่วยเสียงในภาษาเป้าหมาย เช่น ในภาษาไทยไม่มีกลุ่มเสียงพยัญชนะเสียดแทรก โฆษะเลย แต่ในภาษาอังกฤษมีถึง 4 หน่วยเสียง ได้แก่ /v,ð ,z, ʒ/ จึงเป็นปัญหาในการหาอักษรในภาษาไทยที่มีคุณสมบัติทางเสียงใกล้เคียงกันที่สุดมาถอดอักษร

2.3.1.5 การแบ่งพยางค์ในภาษาต้นแบบ เมื่อมีพยัญชนะตัวเดียวอยู่ระหว่างสระ เช่น คำภาษาอังกฤษ money หรือ honey จะถอดพยัญชนะซ้ำสองตัวเพื่อให้อ่านได้สะดวก เป็น มันนีย์ ฮันนีย์ หรือ ถอดอักษรเพียงตัวเดียวตามที่ปรากฏในภาษาอังกฤษโดยให้พยัญชนะตัวนั้นเป็นพยัญชนะท้ายของพยางค์หน้า หรือให้เป็นพยัญชนะต้นของพยางค์ถัดไป เป็น มะนีย์, ฮะนีย์ หรือ มันอีย์, ฮันอีย์

2.3.1.6 อักษรซ้ำในภาษาต้นแบบซึ่งแทนหน่วยเสียงเดียว เช่น ในภาษาอังกฤษ “ll, tt, ss” /l, t, s/ ในการถอดอักษรควรถอดซ้ำหรือไม่

2.3.1.7 ระบบพยัญชนะควบกล้ำในภาษาต้นแบบต่างจากภาษาเป้าหมายมาก เช่น พยัญชนะควบกล้ำในภาษาอังกฤษมีได้ถึง 4 หน่วยเสียงในตำแหน่งท้าย ในการถอดอักษรเป็นไทยโดยเฉพาะในตำแหน่งท้ายจำเป็นต้องใช้เครื่องหมายการันต์เพื่อช่วยให้อ่านได้ แต่มีปัญหาว่าจะใช้เครื่องหมายการันต์อย่างไร

2.3.1.8 พยัญชนะที่ทำหน้าที่ก่อพยางค์ (Syllabic) เช่น /m̩/ ในคำ rhythm /l̩/ ในคำ little ในการถอดอักษรจะใช้สระใดแทน

2.3.2 ปัญหาอื่น ๆในการถอดอักษร

2.3.2.1 ปัญหาอันเนื่องมาจากสำเนียงของภาษาต้นแบบที่ยืมมาต่างกัน เช่น ถ้าภาษาต้นแบบเป็นภาษาอังกฤษมีสำเนียงที่ต่างกัน เช่น สำเนียงอังกฤษและสำเนียงอเมริกัน ตัวอย่างเช่นสำเนียงอังกฤษใช้ /ɑː/ แต่อเมริกันใช้ /æ/ หรือ สำเนียงอังกฤษไม่ออกเสียง /r/ หลังสระ แต่สำเนียงอเมริกันมีการออกเสียง /r/ หลังสระ ในการถอดอักษรจะต้องเลือกว่าจะถอดตามสำเนียงใดสำเนียงหนึ่งให้เหมือนกันทั้งระบบ

2.3.2.2 ปัญหาอันเนื่องมาจากช่วงเวลาของการยืมคำทับศัพท์ คำทับศัพท์บางคำยืมมาเป็นระยะเวลานานแล้ว และถอดอักษรตามลักษณะเสียงในปัจจุบัน จึงมีคำทับศัพท์

จำนวนไม่น้อยที่ไม่เป็นไปตามระบบ เช่น “c” ที่แทน /k/ ในคำยืมเก่าถอดอักษรเป็น “ก” เช่น ในคำ cook ถอดอักษรเป็น กุ๊ก carat ถอดอักษรเป็น กะรัต cap ถอดอักษรเป็น แก๊ป แต่ในคำปัจจุบัน “c” แทน /k/ มักถอดอักษรเป็น “ค” ในตำแหน่งพยัญชนะต้น เช่น condominium ถอดอักษรเป็น คอนโดมิเนียม capsule ถอดอักษรเป็น แคปซูล

หลังจากที่ได้กล่าวถึงปัญหาในการถอดอักษรไปแล้วนั้น ในหัวข้อถัดไปจะกล่าวถึงหลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถาน เนื่องจากวิทยานิพนธ์เล่มนี้ต้องการพัฒนาโปรแกรมการถอดอักษรโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน เพื่อให้การถอดอักษรเป็นไปตามมาตรฐานเดียวกับที่ราชบัณฑิตยสถานได้วางไว้

2.4 หลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถาน

เนื่องจากวิทยานิพนธ์นี้ต้องการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยยึดการถอดอักษรของราชบัณฑิตยสถานเป็นแนวทาง จึงขอนำเสนอหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน (2532) ซึ่งได้วางไว้ดังนี้

1. การทับศัพท์ให้ถอดอักษรในภาษาเดิมพอควรแก่การแสดงที่มาของรูปศัพท์คือเมื่อเห็นคำทับศัพท์แล้วสามารถทราบได้ว่ามาจากคำภาษาอังกฤษคำใด และให้เขียนในรูปที่อ่านได้สะดวกในภาษาไทย
2. การวางหลักเกณฑ์ได้แยกกำหนดหลักเกณฑ์การทับศัพท์ภาษาต่าง ๆ แต่ละภาษาไป เพราะไม่ใช่การถ่ายเสียงแต่เป็นการถอดอักษร
3. คำทับศัพท์ที่ใช้กันมานานจนถือเป็นภาษาไทย และปรากฏในพจนานุกรม ฉบับราชบัณฑิตยสถานแล้ว ให้ใช้ต่อไปตามเดิม เช่น ช็อกโกเลต, ช็อกโกแลต, เชิ้ต, ก๊าซ, แก๊ส ในที่นี้หมายถึงรูปเขียนที่ใช้กันจนติดแล้ว
4. คำวิสามานยนามที่ใช้กันมานานแล้ว อาจใช้ต่อไปตามเดิม เช่น

   Victoria = วิกตอเรีย

   Louis = หลุยส์

   Cologne = โคโลญ
5. ศัพท์วิชาการซึ่งใช้เฉพาะกลุ่ม ไม่ใช่ศัพท์ทั่วไป อาจเพิ่มเติมหลักเกณฑ์ขึ้นตามความจำเป็น

สำหรับหลักเกณฑ์การทับศัพท์ภาษาอังกฤษนั้น ราชบัณฑิตยสถานได้อธิบายไว้ดังนี้

1. สระ ให้ถอดตามการออกเสียงในพจนานุกรมภาษาอังกฤษ โดยเทียบเสียงสระภาษาไทยตามตารางเทียบเสียงสระภาษาอังกฤษ

2. พยัญชนะ ให้ถอดเป็นพยัญชนะภาษาไทยตามหลักเกณฑ์ในตารางเทียบพยัญชนะภาษาอังกฤษ

3. การใช้เครื่องหมายทัณฑฆาต ใช้กำกับพยัญชนะที่ไม่ออกเสียงในภาษาไทยมีรายละเอียดดังนี้

3.1 พยัญชนะตัวที่ไม่ออกเสียงในภาษาไทยเสียงเดียว ให้ใส่เครื่องหมายทัณฑฆาตกำกับไว้ เช่น

horn = ฮอร์น

Windsor = วินด์เซอร์

3.2 คำหรือพยางค์ที่ตัวสะกดมีพยัญชนะที่ไม่ออกเสียงเรียงกันหลายตัว ให้ใส่เครื่องหมายทัณฑฆาต ไว้บนพยัญชนะที่ไม่ออกเสียงตัวสุดท้าย แต่เพียงแห่งเดียว เช่น

Okhotsk = โอค็อตสก์

Barents = แบเร็นตส์

3.3 คำหรือพยางค์ที่มีพยัญชนะไม่ออกเสียงอยู่หน้าเสียงตัวสะกด แล้วยังมีพยัญชนะที่ไม่ออกเสียงตามหลังมาอีก ให้ตัดพยัญชนะที่ไม่ออกเสียงหน้าตัวสะกดออก และใส่เครื่องหมายทัณฑฆาตไว้บนพยัญชนะตัวสุดท้าย เช่น

world = เวิลด์ ในที่นี้ตัดพยัญชนะที่ไม่ออกเสียง “r”

quartz = ควอตซ์ ในที่นี้ตัดพยัญชนะที่ไม่ออกเสียง “r”

Johns = จอนส์ ในที่นี้ตัดพยัญชนะที่ไม่ออกเสียง “h”

first = เฟิสต์ ในที่นี้ตัดพยัญชนะที่ไม่ออกเสียง “r”

4. การใช้ไม้ไต่คู้ ควรใช้ในกรณีต่อไปนี้

4.1 เพื่อให้เห็นแตกต่างจากคำไทย เช่น

log = ล็อก ให้ต่างจาก ลอก

4.2 เพื่อช่วยให้ผู้อ่านแยกพยางค์ได้ถูกต้อง เช่น

Okhotsk = โอค็อตสก์

5. การใช้เครื่องหมายวรรณยุกต์ การเขียนคำทับศัพท์ ไม่ต้องใส่เครื่องหมายวรรณยุกต์ ยกเว้นในกรณีที่คำนั้นมีเสียงซ้ำกับคำไทย จนทำให้เกิดความสับสน อาจใส่เครื่องหมายวรรณยุกต์ได้ เช่น

coke = โค้ก

coma = โคม่า

6. พยัญชนะซ้อน (double letter) คำที่มีพยัญชนะซ้อนเป็นตัวสะกด ถ้าเป็นศัพท์ทั่วไป ให้ตัดออกตัวหนึ่ง เช่น

football = ฟุตบอล

แต่ถ้าเป็นศัพท์ทางวิชาการหรือวิสามานยนามให้เก็บไว้ทั้ง ๒ ตัว โดยใส่เครื่องหมายทัณฑฆาตไว้ที่ตัวท้าย เช่น

cell = เซลล์

James Watt = เจมส์ วัตต์

ถ้าพยัญชนะซ้อนอยู่กลางศัพท์ให้ถือว่า พยัญชนะซ้อนตัวแรกเป็นตัวสะกดของพยางค์หน้า และพยัญชนะซ้อนตัวหลัง เป็นพยัญชนะต้นของพยางค์ต่อไป ฉะนั้น การใช้พยัญชนะตัวสะกดและพยัญชนะต้น จะต่างกันตามหลักเกณฑ์การเทียบพยัญชนะ ในตารางข้างท้าย เช่น

pattern = แพตเทิร์น

Missouri = มิสซูรี

broccoli = บรอกโคลี

7. คำที่ตัวสะกดของพยางค์หน้าออกเสียงเป็นพยัญชนะต้นของพยางค์ตัวต่อไปด้วย ให้ถือหลักเกณฑ์ดังนี้

7.1 ถ้าสระของพยางค์หน้าเป็นเสียงสระอะ ซึ่งเมื่อทับศัพท์ต้องใช้รูปไม้หันอากาศ ให้ซ้อนพยัญชนะตัวสะกด ของพยางค์หน้า อีกตัวหนึ่งเพื่อให้เป็นพยัญชนะต้นของพยางค์ต่อไป เช่น

couple = คัปเปิล

double = ดับเบิล

7.2 ถ้าสระของพยางค์หน้าเป็นสระอื่นที่ไม่ใช่สระอะ ให้ทับศัพท์ตามรูปพยัญชนะภาษาอังกฤษโดยไม่ต้องซ้อนพยัญชนะ เช่น

California = แคลิฟอร์เนีย แทนที่จะเป็น แคลลิฟอร์เนีย

general = เจเนอรัล แทนที่จะเป็น เจนเนอรัล

7.3 ถ้าเป็นคำที่เกิดจากการเติมปัจจัย เช่น -er, -ing, -ic, -y และการทับศัพท์ตามรูปพยัญชนะภาษาอังกฤษดังข้อ ๗.๒ อาจทำให้ออกเสียงผิดไปจากภาษาเดิมมาก ให้ซ้อนพยัญชนะตัวสะกดของพยางค์ต้นอีกหนึ่งเพื่อให้เห็นเค้าคำเดิม เช่น

sweater = สเวตเตอร์

booking = บุกกิง

Snoopy = สนูปปี

8. คำประสมที่มีเครื่องหมายยัติภังค์ (hyphen) ให้ทับศัพท์โดยเขียนติดต่อกันไป เช่น

cross-stitch = ครอสสติตช์

ยกเว้นในกรณีที่เป็นศัพท์ทางวิชาการหรือวิสามานยนาม เช่น

cobalt-60 = โคบอลต์-๖๐

McGraw-Hill = แมกกรอว์-ฮิลล์

9. คำประสมซึ่งในภาษาอังกฤษเขียนแยกกัน เมื่อทับศัพท์ให้เขียนติดกันไป ไม่ต้องแยกคำตามภาษาเดิม เช่น

calcium carbonate = แคลเซียมคาร์บอเนต

night club = ไนต์คลับ

New Guinea = นิวกินี

10. คำคุณศัพท์ที่มาจากคำนามมีการผสมระหว่างการทับศัพท์กับศัพท์บัญญัติ ซึ่งมีปัญหาว่าจะทับศัพท์ในรูปคำนามหรือคำคุณศัพท์นั้น ให้ถือหลักเกณฑ์ดังนี้

10.1 ถ้าคำคุณศัพท์นั้นมีความหมายเหมือนคำนาม หรือหมายความว่า "เป็นของ" หรือ "เป็นเรื่องของ" คำนามนั้น ให้ทับศัพท์ในรูปคำนาม เช่น

hyperbolic curve = (ส่วนโค้ง)ไฮเพอร์โบลา

electronic charge = (ประจุ)อิเล็กตรอน

focal length = (ความยาว)โฟกัส

10.2 ถ้าคำคุณศัพท์นั้นมีความหมายว่า "เกี่ยวข้องกับ" หรือ "เกี่ยวเนื่องจาก" คำนามนั้น ให้ทับศัพท์ในรูปคำนามโดยใช้คำประกอบ เชิง แบบ อย่าง ทาง ชนิด ระบบ ฯลฯ แล้วแต่ความหมาย เช่น

atomic absorption = (การดูดกลืน)โดยอะตอม

electronic power conversion = (การแปลงผันกำลังเชิง)อิเล็กทรอนิกส์

10.3 ในกรณีที่การทับศัพท์ในรูปคำนามตามข้อ 10.1 และข้อ 10.2 ทำให้เกิดความหมายกำกวมหรือคลาดเคลื่อน ให้ทับศัพท์ในรูปคำคุณศัพท์ เช่น

sulfuric acid = กรดซัลฟิวริก

feudal system = ระบบฟิวดัล

metric system = ระบบเมตริก

hyperbolic function = ฟังก์ชันไฮเพอร์โบลิก

11. คำคุณศัพท์ที่มาจากชื่อบุคคล ให้ทับศัพท์ตามชื่อของบุคคลนั้น ๆ โดยใช้คำ เช่น ของ แบบ ระบบ ฯลฯ ประกอบตามความหมาย เช่น

Euclidean geometry = เรขาคณิตระบบยุคลิด

Eulerian function = ฟังก์ชันแบบออยเลอร์

Napierian logarithm = ลอการิทึมแบบเนเปียร์

ยกเว้นในกรณีที่คำคุณศัพท์ที่มาจากชื่อบุคคล เป็นชื่อเฉพาะที่เป็นที่รู้จักกันทั่วไปในแต่ละวงการ ซึ่งอาจสังเกตได้จากการที่ในภาษาอังกฤษไม่ได้ใช้อักษรตัวใหญ่ขึ้นต้น ให้ทับศัพท์ในรูปคำคุณศัพท์ เช่น

abelian group = กลุ่มอาบีเลียน

12. คำคุณศัพท์เกี่ยวกับชนชาติต่าง ๆ ให้ทับศัพท์ในรูปคำนามที่เป็นชื่อประเทศ เช่น

Swedish people = คนสวีเดน

Hungarian dance = ระบำฮังการี

ยกเว้นชื่อที่เคยใช้มานานแล้ว ได้แก่

ประเทศเยอรมนี ใช้ว่า ...เยอรมัน เช่น ภาษาเยอรมัน

ประเทศกรีซ ใช้ว่า ...กรีก เช่น เรือกรีก

ประเทศไอร์แลนด์ ใช้ว่า ...ไอริช เช่น ชาวไอริช

ประเทศเนเธอร์แลนด์ ใช้ว่า ...ฮอลันดา เช่น ชาวฮอลันดา หรือ ...ดัตช์ เช่น ภาษาดัตช์

ประเทศสวิตเซอร์แลนด์ ใช้ว่า ...สวิส เช่น ผ้าสวิส

สหราชอาณาจักรบริเตนใหญ่และไอร์แลนด์เหนือ ใช้ว่า ...อังกฤษ เช่น คนอังกฤษ

สหรัฐอเมริกา ใช้ว่า ...อเมริกัน เช่น รถอเมริกัน

สำหรับสหภาพโซเวียต ซึ่งในภาษาอังกฤษใช้คำคุณศัพท์ ๒ คำ คือ Soviet...และ Russian...ใช้ว่า ...โซเวียต และ ...รัสเซีย เช่น

Soviet Style (of architecture) = (สถาปัตยกรรม) แบบโซเวียต

Russian food = อาหารรัสเซีย

13. การวางตำแหน่งคำคุณศัพท์ในคำทับศัพท์ ให้ถือหลักเกณฑ์ดังนี้

13.1 คำคุณศัพท์ที่ประกอบคำนามที่เป็นภาษาไทย หรือเป็นคำทับศัพท์ แต่ได้ใช้ในภาษาไทยมาจนถือเป็นคำไทยแล้ว ให้วางคำคุณศัพท์ไว้หลังคำนาม เช่น

cosmic ray = (รังสี)คอสมิก

gross ton = (ตัน)กรอส

13.2 ถ้าทั้งคำคุณศัพท์และคำนามเป็นคำทับศัพท์ที่ยังไม่ถือเป็นคำไทย ให้ทับศัพท์ตรงตามศัพท์เดิม เช่น

Arctic Circle = อาร์กติกเซอร์เคิล

adrenal cortex = อะดรีนัลคอร์เทกซ์

13.3 ถ้าต้องการเน้นว่าคำนามนั้นเป็นสิ่งที่มีหลายชนิดและคำคุณศัพท์ที่ประกอบเป็นชนิดหนึ่งของคำนามนั้น อาจทับศัพท์โดยใช้คำประกอบ แบบ ชนิด ระบบ ฯลฯ มาแทรกไว้ระหว่างคำนามกับคำคุณศัพท์ เช่น

normal matrix = เมทริกซ์แบบนอร์แมล

thermosetting plastic = พลาสติกชนิดเทอร์โมเซตติง

14. คำย่อ ให้เขียนชื่อตัวอักษรนั้น ๆ ลงเป็นภาษาไทย ดังนี้

A = เอ B = บี C = ซี

D = ดี E = อี F = เอฟ

G = จี H = เอช I = ไอ

J = เจ K = เค L = แอล

M = เอ็ม N = เอ็น O = โอ

P = พี Q = คิว R = อาร์

S = เอส T = ที U = ยู

V = วี W = ดับเบิลยู X = เอกซ์

Y = วาย Z = แซด

และให้เขียนโดยไม่ต้องใส่จุดและไม่เว้นช่องไฟตามภาษาต้นฉบับ เช่น

BBC = บีบีซี

F.B.I = เอฟบีไอ

DDT = ดีดีที

15. คำทับศัพท์ที่ผูกขึ้นจากตัวย่อ ซึ่งอ่านออกเสียงได้เสมือนคำคำหนึ่ง มิได้ออกเสียงเรียงตัวอักษร ให้เขียนตามเสียงที่ออกและไม่ต้องใส่จุด เช่น

USIS = ยูซิส

UNESCO = ยูเนสโก

ASEAN = อาเซียน

16. ตัวย่อชื่อบุคคล ให้เขียนโดยใส่จุด และเว้นช่องไฟระหว่างชื่อกับนามสกุลตามภาษาต้นฉบับ เช่น

D.N. Smith = ดี.เอ็น. สมิท

G.H.D. Cold = จี.เอช.ดี. โคลด์

ตารางเทียบเสียงสระภาษาอังกฤษ

A - E - I - O - U - Y

สระ	**ใช้**	**ตัวอย่าง**		
A				
a	แอ	badminton	=	แบดมินตัน
	อะ	aluminium	=	อะลูมิเนียม
	อา	Chicago	=	ชิคาโก
	เอ	Asia	=	เอเชีย
	ออ	football	=	ฟุตบอล
ar	แอ	arrow	=	แอร์โรว์
	อา	bar	=	บาร์
	ออ	ward	=	วอร์ด
	เออ	Edward	=	เอดเวิร์ด
are	แอ	mare	=	แมร์
	หมายเหตุ ยกเว้น are (verb to be) ใช้อาร์			
aa	อา	bazaar	=	บาซาร์
	แอ	Aaron	=	แอรอน
ae	อี	Aegean	=	อีเจียน
	แอ	aerosphere	=	แอโรสเฟียร์
	เอ	sundae	=	ซันเด
aea	เอีย	Judaea	=	จูเดีย
aer	เออ	kaersutite	=	เคอร์ซูไทต์

ai	เอ	Spain	=	สเปน
	ไอ	Cairo	=	ไคโร
air	แอ	Bel Air	=	เบลแอร์
ao	เอา	Mindanao	=	มินดาเนา
au	อา	laugh	=	ลาฟ
	ออ	Augusta	=	ออกัสตา
	เอา	Bissau	=	บิสเซา
	โอ	Auger	=	โอเจอร์
	แอ	Laughlin	=	แลฟลิน
aw	ออ	lawernce	=	ลอว์เรนซ์
ay	เอ	Malay	=	มาเลย์
ayr	แอ	Ayrshire	=	แอร์เชอร์
E				
e	อี	Sweden	=	สวีเดน
	เอ	Lebanon	=	เลบานอน
	อิ	electronics	=	อิเล็กทรอนิกส์
	เอะ	Mexico	=	เม็กซิโก
er	เออ	Canberra	=	แคนเบอร์รา
	อา	Clerk	=	คลาร์ก
ere	เอีย	cashmere	=	แคชเมียร์
ea	อี	Guinea	=	กินี
	เอ	Dead Sea	=	เดดซี
	เอีย	Caribbean	=	แคริบเบียน
ear	แอ	Bear	=	แบร์
	เอีย	gear	=	เกียร์
	อา	Heart	=	ฮาร์ต
	เออ	Pearl Harbour	=	เพิร์ลฮาร์เบอร์
eau	โอ	Beaufort	=	โบฟอร์ด
	อิว	Beauty	=	บิวตี

ee	อี	Gr**ee**nwich	=	กรีนิช
eer	เอีย	b**eer**	=	เบียร์
ei	อี	N**ei**t	=	นีล
	ไอ	**Ei**nsteinium	=	ไอน์สไตเนียม
	เอ	B**ei**rut	=	เบรุต
eir	แอ	h**eir**	=	แอร์
	เอีย	P**eir**se	=	เพียร์ส
	เออ	p**eir**ce	=	เพิร์ซ
eo	อี	p**eo**ple	=	พีเพิล
	เอ	L**eo**minster	=	เลมินสเตอร์
	เอีย	Napol**eo**n	=	นะโปเลียน
	เอียว	Born**eo**	=	บอร์เนียว
eou	โอ	S**eou**l	=	โซล
eu	อิว	l**eu**kemia	=	ลิวคีเมีย
	ยุ	**Eu**rope	=	ยุโรป
	ยู	**Eu**phrates	=	ยูเฟรทีส
	เอีย	ol**eu**m	=	โอเลียม
	อู	R**eu**ben	=	รูเบ็น
eur	เออ	f**leur**-de-lis	=	เฟลอร์เดอลีส์
ew	อิว	N**ew** York	=	นิวยอร์ก
	โอ	s**ew**	=	โซว์
	อู	Andr**ew**	=	แอนดรูว์
ey	เอ	Yardl**ey**	=	ยาร์ดเลย์
	อี	k**ey**	=	คีย์
I				
i	อิ	K**i**ng	=	คิง
	อี	sk**i**	=	สกี
	ไอ	L**i**beria	=	ไลบีเรีย
ir	เออ	z**ir**con	=	เซอร์คอน

	อี	Pam**ir**	=	ปามีร์
ire	ไอ	**Ire**land	=	ไอร์แลนด์
	เออ, เอีย	Hampsh**ire**	=	แฮมป์เชอร์, แฮมป์เชียร์
ia	เอีย	Ind**ia**	=	อินเดีย
ie	อี	**rie**beckite	=	รีเบกไกต์
	เอีย	Sov**iet**	=	โซเวียต
	อาย	p**ie**	=	พาย
	ไอ	neckt**ie**	=	เนกไท
ier	เอีย	glac**ier**	=	เกลเชียร์
iew	อิว	v**iew**	=	วิว
*ion	เอียน	Un**ion**	=	ยูเนียน
	อัน	lot**ion**	=	โลชัน
iu	เอีย	alumin**ium**	=	อะลูมิเนียม
O				
o	โอ	Cair**o**	=	ไคโร
	ออ	T**o**m	=	ทอม
	อะ	Washingt**o**n	=	วอชิงตัน
	อู	T**o**day	=	ทูเดย์
or	ออ	c**or**ruption	=	คอร์รัปชัน
	เออ	Winds**or**	=	วินด์เซอร์
ore	ออ	Thomas M**ore**	=	ทอมัส มอร์
	โอ	Ben M**ore**	=	เบนโมร์
oa	ออ	B**roa**dway	=	บรอดเวย์
	โอ	**Oa**kland	=	โอกแลนด์
	อัว	Sam**oa**	=	ซามัว
oar	ออ	b**oar**d	=	บอร์ด
oe	โอ	J**oe**	=	โจ

* ผู้วิจัยเห็นว่า <n> ไม่ใช่สระ จึงไม่ควรรวม ion ไว้เป็นสระ แต่ w, y หรือ r ซึ่งเป็นพยัญชนะกึ่งสระนั้นอาจจัดเป็นรูปสระได้

	อู	Sh**oe**maker	=	ชูเมกเกอร์
	เออ	g**oe**thite	=	เกอไทต์
	อี	Ph**oe**nix	=	ฟีนิกซ์
oer	เออ	**oer**likon	=	เออร์ลิคอน
oi	ออย	thyr**oi**d	=	ไทรอยด์
	อัว	cham**oi**s	=	ชามัวส์
oo	อุ	f**oo**t	=	ฟุต
	อู	w**oo**d	=	วูด
	อะ	Bl**oo**dsworth	=	บลัดส์เวิร์ท
oor	อัว	M**oor**	=	มัวร์
	ออ	d**oor**	=	ดอร์
	โอ	D**oor**n	=	โดร์น
ou	เอา	c**ou**nter	=	เคาน์เตอร์
	อาว	gr**ou**nd	=	กราวนด์
	โอ	B**ou**lder	=	โบลเดอร์
	อะ	thermoc**ou**ple	=	เทอร์โมคัปเปิล
	ออ	G**ou**gh	=	กอฟ
	อุ	s**ou**p	=	ซุป
	อู	Vanc**ou**ver	=	แวนคูเวอร์
our	ออ	b**our**nonite	=	บอร์โนไนต์
	เออ	Melb**our**ne	=	เมลเบิร์น
	อัว	t**our**	=	ทัวร์
	โอ	M**our**ne	=	โมร์น
ow	โอ	b**ow**ling	=	โบว์ลิง
	เอา	C**ow**pens	=	เคาว์เพนส์
	อาว	t**ow**nhouse	=	ทาวน์เฮาส์
	อู	C**ow**per	=	คูว์เปอร์
oy	ออย	Ll**oy**d	=	ลอยด์
U				

u	อะ	Hungary	=	ฮังการี
	อิว	Cuba	=	คิวบา
	อุ	Lilliput	=	ลิลลิพุต
	อู	Kuwait	=	คูเวต
	ยู	Uranium	=	ยูเรเนียม
	อิ	busy	=	บิซี
ur	เออ	hurricane	=	เฮอร์ริเคน
ure	อัว	Sure	=	ชัวร์
	เออ	lecture	=	เลกเชอร์
	เอียว	Pure	=	เพียวร์
ua	อัว	Guadalupe	=	กัวดาลูป
ue	อิว	Tuesday	=	ทิวส์เดย์
	อู	wuestite	=	วูสไทต์
ui	อุ	fruit	=	ฟรุต
	อู	juice	=	จูซ
	อิ	circuit	=	เซอร์คิต
	ไอ	Ruislip Northwood	=	ไรสลิปนอร์ทวูด
uir	อิว	Muir	=	มิวร์
uy	ไอ	Schuyler	=	สไกลเลอร์
	อาย	Guy	=	กาย
Y				
y	อิ	Odyssey	=	โอดิสซีย์
	อี	Syria	=	ซีเรีย
	ไอ	cyclone	=	ไซโคลน
ye	ไอ	rye	=	ไรย์
yr	เออ	Myrna	=	เมอร์นา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางเทียบพยัญชนะภาษาอังกฤษ

พยัญชนะ		**พยัญชนะต้น**				**ตัวสะกดและตัวการันต์**		
ภาษาอังกฤษ	**ใช้**	**ตัวอย่าง**			**ใช้**	**ตัวอย่าง**		
b	บ	**b**ase	=	เบส	บ	Gi**bb**	=	กิบบ์
c+a	ค	**ca**t	=	แคต	ก	cubi**c**	=	คิวบิก
+o		**co**ne	=	โคน				
+u		**Cu**ba	=	คิวบา				
+r		**cr**own	=	คราวน์				
+l		**Cl**eo	=	คลีโอ				
c+e	ซ	**ce**ll	=	เซลล์	ซ	Gree**ce**	=	กรีซ
+i		**ci**gar	=	ซิการ์				
+y		**cy**clone	=	ไซโคลน				
c (ออกเสียง ซ)	ซ	gla**ci**er	=	เกลเซียร์	-	-		
	หมายเหตุ คำที่พยางค์สุดท้ายเป็น ca, co และ cer (ที่ออกเสียงเกอร์) ซึ่งไทยเรานิยมใช้เสียง ก ให้ใช้ ก							
	ตัวอย่าง							
		Ameri**ca**	=	อเมริกา				
		dis**co**	=	ดิสโก				
		soc**cer**	=	ซอกเกอร์				
ch (ออกเสียง ช)	ช	**Ch**icago	=	ชิคาโก	ช	Bea**ch**	=	บีช
ch (ออกเสียง ค)	ค	**Ch**ios	=	คิออส	ก	Angio**ch**	=	แอนติออก
-ck					ก	Brunswi**ck**	=	บรันสวิก
	หมายเหตุ ck เมื่อเป็นตัวสะกดและพยัญชนะต้นของพยางค์ต่อไปให้ใช้ กก เช่น							
		Ro**ck**y	=	รอกกี				
		lo**ck**et	=	ล็อกเกต				
d	ด	**d**extrin	=	เดกซ์ทริน	ด	Dead Sea	=	เดดซี

f	ฟ	Fox	=	ฟอกซ์	ฟ	Clifion	=	คลิฟตัน
g+e	จ	**ge**stagen	=	เจสตาเจน	จ	rou**ge**	=	รูจ
+i		en**gi**neer	=	เอนจิเนียร์				
+y		**gy**ro	=	ไจโร				
g+a	ก	**ga**laxy	=	กาแล็กซี	ก	ma**g**nesium	=	แมกนีเซียม
+e		for**ge**t-me-not	=	ฟอร์เกตมีนอต				
+i		**gi**ft	=	กิฟต์				
+o		**go**lf	=	กอล์ฟ				
+u		**gu**lf	=	กัลฟ์				
+l		**Gl**adstone	=	แกลดสโตน				
+r		**gr**and	=	แกรนด์				
gh	ก	**gh**etto	=	เกตโต	ฟ	Dou**gh**	=	กอฟ
					ก	Pittsbur**gh**	=	พิตส์เบิร์ก
gh(ไม่ออกเสียง)	-				-	Hu**gh**	=	ฮิว
gn (g ไม่ออกเสียง)	น	**gn**eiss	=	ไนส์	น	desi**gn**	=	ดีไซน์
	หมายเหตุ gn ถ้าเป็นตัวสะกดและพยัญชนะต้นของพยางค์ถัดไป แล้วออกเสียง /ñ/ ใช้ นญ เช่น							
		Bolo**gn**a	=	โบโลนญา				
		Co**gn**ac	=	คอนญัก				
gu (ออกเสียง กว)	กว	pen**gu**in	=	เพนกวิน				
gu (ออกเสียง ก)	ก	**Gu**ildford	=	กิลด์ฟอร์ด	ก	lea**gu**e	=	ลีก
h	ฮ	**H**aematite	=	ฮีมาไทต์				
h (ไม่ออกเสียง)	-	**h**onour	=	ออเนอร์	ห	Jo**h**n	=	จอห์น
j	จ	**J**im	=	จิม				

k	ค	**K**ansas	=	แคนซัส	ก	Yor**k**	=	ยอร์ก
k (เมื่อเป็น	ก	bun**k**er	=	บังเกอร์				
พยัญชนะต้น ของ		mar**k**et	=	มาร์เกต				
พยางค์ สุดท้าย)		Yan**k**ee	=	แยงกี				
kh	ค	**kh**artoum	=	คาร์ทูม	ก	Si**kh**	=	ซิก
l	ล	**l**ocket	=	ล็อกเกต	ล	She**ll**	=	เชลล์
m	ม	**m**icro	=	ไมโคร	ม	To**m**	=	ทอม
n	น	**n**ucleus	=	นิวเคลียส	น	cyclo**n**e	=	ไซโคลน
	หมายเหตุ n เมื่อเป็นตัวสะกดและมีพยัญชนะ c ch g k qu ฯลฯ ตามแล้วทำให้เสียง n ที่เป็นตัวสะกดออกเสียงเป็น ง ให้ถอด n เป็น ง เช่น							
		A**n**glo-Saxon	=	แองโกลแซกซัน				
		fu**n**ction	=	พังก์ชัน				
		pare**n**chyma	=	พาเรงคิมา				
		Fra**n**k	=	แฟรงก์				
p	พ	**p**araola	=	พาราโบลา	ป	ca**p**sule	=	แคปซูล
	หมายเหตุ p เมื่อเป็นพยัญชนะต้นให้ใช้ พ โดยตลอด ยกเว้นกลุ่มพยัญชนะบางกลุ่มที่ไทย นิยมใช้เสียง ป ให้ใช้ ป ดังนี้ super-, -pa, -pean, -per, -pia, -pic, -ping, -pion, -po, -pus และ -py							
	ตัวอย่าง							
		superman	=	ซูเปอร์แมน				
		Euro**pa**	=	ยูโรปา				
		bum**per**	=	บัมเปอร์				
		to**pic**	=	ทอปปิก				
		shop**ping**	=	ชอปปิง				
		hip**py**	=	ฮิปปี				
		hip**po**	=	ฮิปโป				
		olym**pus**	=	โอลิมปัส				

ph	ฟ	**ph**osphorous	=	ฟอสฟอรัส	ฟ	gra**ph**	=	กราฟ
q	ก	**Q**atar	=	กาตาร์	ก	Ira**q**	=	อิรัก
qu (ออกเสียง ควฺ)	ควฺ	**Qu**ebec	=	ควิเบก	-	-		
qu (ออกเสียง ค)	ค	Li**qu**or	=	ลิเคอร์	ก	Mozambi**que**	=	โมซัมบิก
r	ร	**r**adium	=	เรเดียม	ร	bar**l**ey	=	บาร์เลย์
rh	ร	**rh**odonite	=	โรโดไนต์	ห	myr**rh**	=	เมอรห์
		mur**rh**a	=	เมอร์รา				
s (+ สระ)	ซ	**s**ilicon	=	ซิลิคอน	ส	Lago**s**	=	ลากอส
(+ พยัญชนะ)	ส	**S**weden	=	สวีเดน	-	-		
(ออกเสียง ซ)	ซ	A**s**ia	=	เอเชีย	-	-		
son (อยู่ท้ายชื่อ)	สัน	John**son**	=	จอห์นสัน				
's	-	-			ส์	King**'s** Cup	=	คิงส์คัป
sc (ออกเสียง ซ)	ซ	**sc**ene	=	ซีน				
(ออกเสียง สก)	สก	**sc**rew	=	สกรูว	สก	di**sc**	=	ดิสก์
sch (ออกเสียง ซ)	ซ	**sch**eelite	=	ซีไลต์				
(ออกเสียง ช)	ช	**sch**ism	=	ชิซึม				
(ออกเสียง สก)	สก	**sch**ool	=	สกูล				
sh	ช	**sh**amal	=	ชามาล	ช	har**sh**	=	ฮาร์ช
sk	สก	**sk**yros	=	สกิรอส	สก	ta**sk**	=	ทาสก์
sm	-	-			ซึม	protopla**sm**	=	โพรโทพลาซึม

sp	สป	**sp**ray	=	สเปรย์				
		spore	=	สปอร์				
st	สต	**St**anford	=	สแตนฟอร์ด				
		strip	=	สตริป				
	หมายเหตุ sc sk sp และ st ที่ถอดเป็น สก, สก, สป, และ สต ตามลำดับ ถ้า s อยู่กลางศัพท์ และเป็นตัวสะกดของพยางค์หน้า ให้ถอด c k p และ t นั้น เป็น ค ค พ และ ท เช่น							
		Wi**sc**onsin	=	วิสคอนซิน				
		Mu**sk**egon	=	มัสคีกัน				
		a**sp**aragus	=	แอสพารากัส				
		di**st**emper	=	ดิสเทมเปอร์				
t	ท	**T**asmania	=	แทสเมเนีย	ต	Kuwai**t**	=	คูเวต
		trombone	=	ทรอมโบน				
	หมายเหตุ t เมื่อเป็นพยัญชนะต้น ใช้ ท โดยตลอด ยกเว้นกลุ่มพยัญชนะบางกลุ่มที่ไทยเรานิยมใช้เสียง ต ให้ใช้ ต ดังนี้ anti-, auto-, inter-, multi-, photo-; ta, -ter, -ti, -tic, -ting, -tis, -to, -ton, - tor, -tre, - tum, -tus และ -ty เช่น							
		antibody	=	แอนติบอดี				
		intercom	=	อินเตอร์คอม				
		compu**ter**	=	คอมพิวเตอร์				
		quan**tum**	=	ควอนตัม				
		zygoma**ta**	=	ไซโกมาตา				
th	ท	**th**orium	=	ทอเรียม	ท	zeni**th**	=	เซนิท
thm	-	-			ทึม	biorhy**thm**	=	ไบโอริทึม
						logari**thm**	=	ลอการิทึม
ti (ออกเสียง ช)	ช	na**ti**on	=	เนชัน	-			
		stron**ti**um	=	สตรอนเชียม				
v	ว	**v**olt	=	โวลต์	ฟ	lo**v**e	=	เลิฟ
(เมื่อเป็นตัวสะกด						perov**s**kite	=	เพอรอฟสไกต์

ของพยางค์ต้นและ								
เป็นตัวนำของพยางค์								
ต่อไปด้วย)	-	-			ฟว	Livingstone	=	ลิฟวิงสโตน
w	ว	**W**ales	=	เวลส์	ว	co**w**boy	=	คาวบอย
wh (ออกเสียง ว)	ว	**Wh**ite	=	ไวต์	-	-		
wh (ออกเสียง ฮ)	ฮ	**Wh**ewell	=	ฮิวเอลล์				
x	ซ	**X**enon	=	ซีนอน	กซ	bo**x**er	=	บอกเซอร์
						o**x**ford	=	ออกซฟอร์ด
						ony**x**	=	โอนิกซ์
y	ย	**Y**ale	=	เยล	ย	ke**y**	=	คีย์
z	ซ	**z**one	=	โซน	ซ	Vadu**z**	=	วาดุซ
	หมายเหตุ							
		๑. คำหรือพยางค์ที่มีพยัญชนะต้นหลายตัว และตัวหน้าไม่ออกเสียง เมื่อเขียนทับศัพท์เป็นภาษาไทย ไม่ต้องใส่พยัญชนะตัวที่ไม่ออกเสียง เช่น						
		gnat	=	แนต				
		knight	=	ไนต์				
		psycho	=	ไซโค				
		pneumonia	=	นิวมอเนีย				
		๒. วิสามานยนาม ที่ในภาษาอังกฤษออกเสียงเฉพาะพิเศษนอกเหนือจากที่กำหนดไว้ในตาราง ให้ถอดตามการออกเสียง เช่น						
		Worcester	=	วูสเตอร์				
		Marble Arch	=	มาร์บะลาช				

สำหรับวิทยานิพนธ์เล่มนี้จะไม่ได้พัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยให้ครอบคลุมหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถานทั้งหมดดังที่ได้กล่าวไปข้างต้น โดยที่จะครอบคลุมเพียงในส่วนของการถอดอักษรภาษาอังกฤษเป็นภาษาไทยทั้งคำ โดยที่

ไม่มีการแปลคำอังกฤษเป็นคำไทยในบางส่วน เช่น ในกรณีของคำว่า Hungarian dance ที่ต้องเขียนเป็น “ระบำฮังการี” สามารถดูวิธีดำเนินการวิจัยโดยละเอียดได้ในบทที่ 3

หลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานข้างต้นนี้จะเห็นได้ว่าถึงแม้ว่าราชบัณฑิตยสถานได้ให้บริบทของการถอดอักษรส่วนใหญ่ไว้แล้วนั้น แต่อักษรบางตัวโดยเฉพาะสระในภาษาอังกฤษนั้นราชบัณฑิตยสถานไม่ได้บอกบริบทการถอดอักษรเอาไว้ เช่น a, i และ o ดังสามารถดูได้จากตารางด้านบน จึงทำให้คนทั่วไปเกิดความสับสนในการถอดอักษรได้ ดังนั้นเมื่อจะนำหลักเกณฑ์ดังกล่าวมาเขียนเป็นกฎให้คอมพิวเตอร์เข้าใจได้จึงจำเป็นต้องหาบริบทการถอดอักษรโดยละเอียด เพื่อที่จะให้ได้ผลลัพธ์ที่ถูกต้องและแม่นยำออกมา ในการเขียนกฎให้กับคอมพิวเตอร์ประมวลผลข้อมูลและทำการถอดอักษรนั้น ได้มีผู้เคยทำการเขียนโปรแกรมไว้แล้วในการถอดอักษรและการถ่ายเสียงในหลายภาษา ซึ่งทั้งการถอดอักษรและการถ่ายเสียงนั้นมีลักษณะที่ร่วมกันคือเป็นการแปลงสัญลักษณ์ตัวอักษรชุดหนึ่ง (รูปเขียน) ไปเป็นสัญลักษณ์ในอีกชุดหนึ่ง (รูปเขียนหรือรูปแทนเสียง) ผู้วิจัยจึงได้ทบทวนวรรณกรรมและได้นำมาเป็นแนวทางในการเขียนกฎให้คอมพิวเตอร์เข้าใจ โดยขั้นตอนแรกนั้นจะขออธิบายเกี่ยวกับแนวทางในการแปลงสัญลักษณ์ชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่ง จากข้อมูลที่เคยผู้ทำมาแล้วนั้น สามารถศึกษาได้จากการแปลงอักษรเป็นอักษรและจากการแปลงอักษรเป็นรูปแทนเสียง ดังต่อไปนี้

2.5 แนวทางในการใช้คอมพิวเตอร์แปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่ง

สำหรับแนวทางในการแปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่งนั้น เมื่อจำแนกตามเกณฑ์ที่ว่าการแปลงสัญลักษณ์นั้นต้องอาศัยตัวกลางหรือไม่ จะสามารถแบ่งออกได้เป็น2 แนวใหญ่ๆ ดังนี้

2.5.1 การใช้คอมพิวเตอร์แปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่งโดยตรง

การใช้วิธีการแปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่งโดยตรงนี้แบ่งเป็น 2 ลักษณะ คือ

2.5.1.1 **การถอดอักษร** การถอดอักษรนั้นคือการแปลงอักษรจากภาษาหนึ่งไปเป็นอักษรอีกภาษาหนึ่งซึ่งในการใช้คอมพิวเตอร์แปลงอักษรโดยตรงนั้นจะทำการถอดอักษรใน

ภาษาหนึ่งไปเป็นอักษรในอีกภาษาหนึ่งโดยไม่ผ่านตัวกลาง ตัวอย่างเช่น งานของ Charoenporn, Chotimongkol, and Sornlertlamvanich (1999) ซึ่งศึกษาเรื่องการแปลงอักษรภาษาไทยเป็นอักษรโรมันโดยใช้คอมพิวเตอร์ โดยใช้แนวทางการแปลงตัวอักษรไทยเป็นตัวอักษรโรมันโดยตรง เพราะงานวิจัยนี้มองว่าปัญหาในการแปลงอักษรไทยเป็นอักษรโรมันนั้นอยู่ที่การแยกพยางค์ ปัญหานี้แก้ได้โดยอาศัยกฎทางภาษาศาสตร์ช่วยในการศึกษาเรื่องการออกเสียงออกการตัดพยางค์ เมื่อสามารถตัดพยางค์ได้แล้ว ก็สามารถแปลงอักษรไทยเป็นอักษรโรมันได้โดยตรง

2.5.1.2 **การถ่ายเสียง** การถ่ายเสียงนั้นคือการแปลงรูปอักษรไปเป็นรูปแทนเสียงของภาษาเดียวกัน ซึ่งจะทำได้โดยการแปลงอักษรไปเป็นรูปแทนเสียงได้โดยตรง ตัวอย่าง เช่น งานของ Van den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่อง การแปลงรูปอักษรเป็นรูปแทนเสียงโดยใช้คลังข้อมูลของคำที่ถ่ายเสียงแล้ว ได้ใช้แนวทางการแปลงจากรูปอักษรเป็นรูปแทนเสียงโดยตรงโดยไม่ผ่านตัวกลาง เพราะผู้วิจัยเห็นว่าการสร้างกฎการถ่ายเสียงจากคลังข้อมูลนั้นสามารถให้ผลที่แม่นยำแล้ว

2.5.2 การใช้คอมพิวเตอร์แปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่ง โดยผ่านตัวกลาง

การแปลงสัญลักษณ์จากชุดหนึ่งไปเป็นสัญลักษณ์อีกชุดหนึ่งโดยผ่านตัวกลางนั้นคือ การที่ต้องอาศัยตัวกลางในการแปลงสัญลักษณ์ชุดหนึ่งไปเป็นอีกชุดหนึ่งด้วย โดยแนวนี้จะพบเป็นส่วนใหญ่ของการศึกษาที่ผ่านมา การแปลงสัญลักษณ์โดยผ่านตัวกลางนั้นพบว่ามีการใช้ตัวกลางที่แตกต่างกันนั่นคือ ส่วนใหญ่จะใช้รูปแทนเสียง (phoneme) เป็นตัวกลางทำได้โดย

1) เขียนคำในภาษาดั้งเดิม
2) แปลงจากรูปอักษรในภาษาดั้งเดิม ไปเป็นรูปแทนเสียงของภาษาดั้งเดิมก่อน
3) เปรียบเทียบเสียงของภาษาดั้งเดิมกับภาษาเป้าหมาย แล้วจึงแปลงไปเป็นรูปแทนเสียงของภาษาเป้าหมาย
4) แปลงจากรูปแทนเสียงของภาษาเป้าหมายไปเป็นรูปอักษรของภาษาเป้าหมาย
5) เขียนคำในภาษาเป้าหมาย

จะเห็นได้ว่าการใช้คอมพิวเตอร์แปลงสัญลักษณ์โดยผ่านตัวกลางนี้ต้องใช้ขั้นตอนซับซ้อนถึง 5 ขั้นตอน เพื่อให้ได้มาซึ่งการถอดอักษรเป็นภาษาเป้าหมาย ตัวอย่างของงานที่แปลง

สัญลักษณ์โดยผ่านตัวกลางนี้ ได้แก่ งานของ Knight and Graehl (1997) ซึ่งศึกษาเรื่อง การถอดอักษรคาตาคานะในภาษาญี่ปุ่นที่เขียนทับศัพท์ภาษาอังกฤษกลับไปเป็นภาษาอังกฤษโดยใช้เครื่องคอมพิวเตอร์ ได้ใช้แนวทางการถอดอักษรโดยใช้ตัวกลาง เพราะในการถอดอักษรญี่ปุ่นที่เขียนทับศัพท์กลับไปเป็นภาษาอังกฤษนั้นจำเป็นต้องใช้ตัวกลางคือรูปแทนเสียง แล้วจึงถอดกลับไปเป็นอักษรภาษาอังกฤษได้ นอกจากนี้งานของ Stalls and Knight (1998)ซึ่งได้ศึกษาเรื่องถอดอักษรภาษาอาหรับที่เขียนทับศัพท์กลับไปเป็นภาษาอังกฤษ ได้ใช้แนวทางการถอดอักษรโดยใช้ตัวกลางคือรูปแทนเสียงเช่นเดียวกัน การที่งานเหล่านี้เลือกที่จะแปลงสัญลักษณ์โดยผ่านตัวกลางทั้งๆที่มีขั้นตอนซับซ้อนกว่า เพราะเชื่อว่าการแปลงสัญลักษณ์ผ่านตัวกลางนี้จะได้ผลการถอดอักษรที่แม่นยำ มีความผิดพลาดน้อยกว่าการแปลงสัญลักษณ์โดยตรง

นอกจากนี้บางคนก็ใช้รหัสที่ตั้งขึ้นมาเป็นตัวกลางในการถอดอักษร เช่น มีการนำ Soundex ซึ่งคือการเปลี่ยนค่าตัวอักษรเป็นรหัสตัวเลข มาเป็นตัวกลาง เช่น ในงานของ Suwanvisat and Prasitjutrakul (1998) ซึ่งได้ศึกษาเรื่อง การค้นคืนข้ามภาษาไทยและอังกฤษ ได้ใช้แนวทางการถอดอักษรโดยใช้ตัวกลางเป็น Soundex เพราะ Soundex นั้นสามารถโยงไปถึงภาษาดั้งเดิมและภาษาเป้าหมาย ทำให้สามารถค้นคืนข้ามภาษาได้ จะสังเกตเห็นว่างานที่ใช้ soundex นี้จะแตกต่างจากงานของ Knight and Graehl (1997) ตรงที่งานของ Knight and Graehl (1997) นั้นเป็นการแปลงจากตัวอักษรดั้งเดิม ไปเป็นตัวกลางนั่นคือ รูปแทนเสียง แล้วจึงแปลงไปเป็นตัวอักษรเป้าหมาย ส่วนงานของ Suwanvisat and Prasitjutrakul (1998) นั้นใช้วิธีแปลงอักษรในภาษาหนึ่งไปเป็นตัวกลางนั่นคือ ตัวเลข และแปลงตัวอักษรอีกภาษาให้เป็นตัวเลขเช่นเดียวกัน เพื่อที่จะได้ค้นคืนข้ามภาษาได้

จะเห็นได้ว่าจากงานที่ได้ศึกษาไปนั้นจะรวมงานที่เกี่ยวกับการค้นคืนข้ามภาษา (cross-language retrieval) ไปด้วยหลายงาน งานเหล่านี้จะใช้การถอดอักษรมาเป็นส่วนช่วยในการค้นคืนข้ามภาษา โดยจะมีจุดประสงค์ต่างไปจากงานเกี่ยวกับการถอดอักษรโดยตรง นั่นคือการค้นคืนข้ามภาษานั้นจะมองว่าเมื่อใส่ข้อมูลเข้าไปแล้ว จะต้องได้ผลลัพธ์ซึ่งเป็นตัวแปรของคำทั้งหมดออกมา เพื่อที่ว่าจากตัวแปรเหล่านี้จะไปกำหนดให้คอมพิวเตอร์จำให้ได้ว่ามาจากคำเดียวกัน และสามารถจัดการกับผลลัพธ์หลายๆตัวนี้ได้ และจะได้ค้นคืนข้ามภาษาเจอคำในเอกสารที่ต้องการ แต่ในกรณีของการถอดอักษรนั้น เมื่อใส่คำเข้าไป 1 คำ จะสามารถกำหนดให้มีผลลัพธ์ตัวเดียวหรือหลายตัวก็ได้ โดยปกติในการถอดอักษรจริงๆ เรามักถอดมาไม่ตรงกัน จึงมี

ผลลัพธ์หลายตัว แต่สำหรับวิทยานิพนธ์ที่จะทำนี้คาดหวังว่าเมื่อใส่ข้อมูลเข้าไป1 คำ จะได้ผลลัพธ์ออกมา 1 คำที่เป็นมาตรฐานเดียวกันตามราชบัณฑิตยสถาน

2.6 แนวทางในการกำหนดกฎในการถอดอักษรให้คอมพิวเตอร์เรียนรู้

เมื่อเราได้ศึกษาแนวทางการแปลงสัญลักษณ์จากชุดหนึ่งไปเป็นอีกชุดหนึ่งไปแล้วนั้น สิ่งที่ต้องพิจารณาต่อไปคือแนวทางในการกำหนดกฎการถอดอักษรเพื่อให้คอมพิวเตอร์เข้าใจได้ กฎในการถอดอักษรสำหรับคอมพิวเตอร์นี้จำเป็นต้องมีบริบทการถอดอักษรภาษาหนึ่งไปเป็นอีกภาษาหนึ่งอย่างละเอียดและเป็นขั้นตอน เพื่อที่จะให้ผลลัพธ์ที่ออกมาถูกต้องและแม่นยำ ในส่วนของกฎการถอดอักษรสำหรับคอมพิวเตอร์นั้นต่างจากกฎที่ราชบัณฑิตยสถานให้ไว้ตรงที่กฎของคอมพิวเตอร์จะละเอียดกว่า จะต้องมีการบอกบริบทหน้าและหลังสำหรับอักษรทุกตัว ซึ่งต่างจากกฎของราชบัณฑิตยสถานที่ในบางตัวอักษรไม่ได้บอกบริบทที่ละเอียดไว้ โดยเฉพาะสระในภาษาอังกฤษดังที่กล่าวไปแล้วข้างต้น กฎการถอดอักษรของคอมพิวเตอร์นั้นเปรียบเสมือนคำสั่งที่ทำให้คอมพิวเตอร์ทำงานให้ถูกต้องตามที่เราต้องการ หากไม่มีกฎ คอมพิวเตอร์ก็จะไม่สามารถทำงานได้หรืออาจจะได้ผลลัพธ์ที่ไม่ถูกต้องแม่นยำออกมา จากการศึกษาที่ผ่านมานั้นสามารถแบ่งวิธีการในการกำหนดกฎในการถอดอักษรของคอมพิวเตอร์ได้ดังนี้

2.6.1 การใช้คนกำหนดกฎให้คอมพิวเตอร์ (Manual construction)

การกำหนดกฎเองนั้น คือ การที่คนเป็นคนกำหนดกฎให้คอมพิวเตอร์ได้เรียนรู้เอง โดยคนเป็นคนกำหนดกฎว่าจะแปลงอักษรนี้เป็นอักษรอีกตัวหนึ่งในบริบทใด ตัวอย่างเช่นงานของ Kawtrakul et al (1998) ซึ่งได้ศึกษาเรื่องการถอดอักษรภาษาไทยที่เขียนทับศัพท์กลับเป็นภาษาอังกฤษเพื่อใช้ในการค้นคืนข้ามภาษาในเอกสารภาษาไทย ได้ใช้แนวทางการกำหนดกฎให้คอมพิวเตอร์จำลองการถอดอักษรเป็นภาษาอังกฤษ (Backward Transliteration Model) มาช่วยในการแยกพยางค์ การถ่ายเสียงและการหาคำภาษาอังกฤษที่มีปัญหาในการจับคู่คำกับภาษาไทย ส่วนที่เป็นกฎที่กำหนดด้วยคนคือ กฎการถ่ายเสียง

2.6.2 การใช้คลังข้อมูลในการกำหนดกฎให้คอมพิวเตอร์ (Corpus-based)

การใช้คลังข้อมูลเพื่อมาช่วยในการกำหนดกฎให้คอมพิวเตอร์นั้น ทำได้โดยการเก็บรวบรวมคลังข้อมูล (corpus) เกี่ยวกับข้อมูลที่ต้องการนำมาพัฒนาเป็นโปรแกรมให้ได้จำนวนมากพอสมควร จากนั้นเขียนโปรแกรมให้คอมพิวเตอร์ได้เรียนรู้จากคลังข้อมูลนั้นๆ แล้วใช้ผลการเรียนรู้

สร้างเป็นกฎขึ้นมาจากคลังข้อมูลนั้นๆ ตัวอย่างเช่น งานของ Van den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่องการแปลงรูปอักษรเป็นรูปแทนเสียงโดยการใช้คลังข้อมูลของคำที่ถ่ายเสียงแล้วมาช่วยในการสร้างกฎ โดยการเขียนโปรแกรมให้คอมพิวเตอร์เรียนรู้และสร้างกฎขึ้นมาเองจากคลังข้อมูลนั้น

2.7 **การจับคู่ข้อมูล** (Alignment)

หลังจากศึกษาแนวทางในการกำหนดกฎให้คอมพิวเตอร์เรียนรู้แล้วนั้น จะเห็นได้ว่าหากต้องการกำหนดกฎให้คอมพิวเตอร์โดยใช้คลังข้อมูล ก็จำเป็นต้องศึกษาเรื่องการเตรียมข้อมูลที่จะนำไปรวบรวมเป็นคลังข้อมูล เพื่อที่ข้อมูลเหล่านี้จะสามารถนำมาสร้างเป็นกฎต่อไปได้ โดยที่การเตรียมข้อมูลนั้น จะต้องมีการนำข้อมูลของภาษาดั้งเดิมและภาษาเป้าหมาย หรือรูปอักษรและ/หรือรูปแทนเสียงที่จะนำมาใช้นั้นมาจับคู่กันก่อน กระบวนการนี้เรียกว่า “การจับคู่ข้อมูล”

การจับคู่ข้อมูลนั้นคือการนำชุดข้อมูลที่ต้องการศึกษามาจับคู่กัน ในกรณีของการถอดอักษรนั้นจะจับคู่ตัวอักษรในภาษาหนึ่งกับตัวอักษรในอีกภาษาหนึ่งที่มีลักษณะทางเสียงคล้ายคลึงกัน ส่วนในกรณีการถ่ายเสียง จะเป็นการจับคู่รูปอักษรกับรูปแทนเสียงที่สอดคล้องกัน ข้อมูลที่จับคู่กันแล้วทั้งหมดนั้นจะนำไปเป็นข้อมูลที่ให้คอมพิวเตอร์ได้เรียนรู้ และสร้างเป็นกฎการถอดอักษรหรือการถ่ายเสียงออกมา จากการศึกษาที่ผ่านมานั้นพบการจับคู่กันระหว่างรูปอักษรและรูปแทนเสียง เช่น ในงานของ Van den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่องการแปลงรูปอักษรเป็นรูปแทนเสียงโดยการใช้คลังข้อมูล ได้ใช้แนวทางการจับคู่ข้อมูลรูปอักษรและรูปแทนเสียงในคลังข้อมูล นอกจากนี้ และในงานของ Kang and Kim (2000) ซึ่งได้ศึกษาเรื่องการถอดอักษรภาษาอังกฤษเป็นภาษาเกาหลีและการถอดอักษรภาษาเกาหลีที่เขียนทับศัพท์กลับเป็นภาษาอังกฤษโดยใช้รูปแทนเสียง ได้ใช้แนวทางการจับคู่สระภาษาเกาหลีกับสระภาษาอังกฤษ และจับคู่พยัญชนะภาษาเกาหลีกับพยัญชนะภาษาอังกฤษ ในการจับคู่ข้อมูลนั้นสามารถทำได้โดยวิธีต่อไปนี้

2.7.1 **การใช้คนจับคู่ข้อมูล** Manual Alignment

การจับคู่ข้อมูลเองนั้น ทำได้โดยการจับคู่ข้อมูลชุดหนึ่งกับอีกชุดหนึ่งโดยใช้มนุษย์เป็นผู้ทำ ซึ่งอาจจะเป็นการจับคู่ตัวอักษรกับรูปแทนเสียง หรือจับคู่ตัวอักษรกับตัวอักษรเลยก็ได้ โดยวิธีนี้จะมีความถูกต้องและแม่นยำค่อนข้างสูงแต่ต้องใช้เวลามาก

2.7.2 **การจับคู่ข้อมูลอัตโนมัติ** Automatic Alignment

การจับคู่ข้อมูลอัตโนมัติ คือ การจับคู่ข้อมูลชุดหนึ่งกับอีกชุดหนึ่งโดยการเขียนวิธีการคิดคำนวณ (algorithm)ให้คอมพิวเตอร์จับคู่ชุดข้อมูลให้ โดยวิธีการคิดคำนวณนั้นเป็นวิธีคิดให้กับคอมพิวเตอร์เข้าใจว่าจะสามารถจับคู่ชุดข้อมูลอย่างไร ในส่วนของวิธีการคิดคำนวณที่ใช้จับคู่ข้อมูลนี้เราสามารถใช้จากของที่มีผู้ทำไว้แล้ว หรือจะเขียนขึ้นใหม่เองก็ได้ จากการศึกษาที่ผ่านมา เช่น ในงานของ Van den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่องการแปลงรูปอักษรเป็นรูปแทนเสียงโดยการใช้คลังข้อมูล ได้ใช้แนวทางการจับคู่ข้อมูลอัตโนมัติ โดยใช้วิธีการคิดคำนวณที่เรียกว่า EM algorithm ในการจับคู่โดยอัตโนมัติ ผลที่ได้ออกมานั้นจะได้คู่ของชุดข้อมูลของ 2 ภาษาออกมาเป็นคู่ๆ การจับคู่ข้อมูลแบบอัตโนมัติที่ใช้คอมพิวเตอร์จับคู่ให้นี้จะสะดวก ใช้เวลาและค่าใช้จ่ายน้อยกว่าใช้คนทำเอง

2.7.3 **การจับคู่ข้อมูลโดยใช้วิธีเฉพาะ** Heuristic Method

การจับคู่ข้อมูลโดยใช้วิธีเฉพาะ คือ ผู้เขียนโปรแกรมใช้วิธีเฉพาะที่คิดขึ้นมาเองเป็นวิธีในการจับคู่ข้อมูล เช่นในงานของ Kang and Kim (2000) ซึ่งได้ศึกษาเรื่องการถอดอักษรภาษาอังกฤษเป็นภาษาเกาหลีและการถอดอักษรภาษาเกาหลีที่เขียนทับศัพท์กลับเป็นภาษาอังกฤษโดยใช้รูปแทนเสียง ได้ใช้แนวทางการจับคู่สระภาษาเกาหลีกับสระภาษาอังกฤษก่อนแล้วจึงจับคู่พยัญชนะภาษาเกาหลีกับพยัญชนะภาษาอังกฤษซึ่งเป็นวิธีที่ผู้วิจัยคิดขึ้นมาเฉพาะเพื่อการวิจัยนี้

2.8 **แบบจำลองที่ใช้** (Model)

เมื่อได้ศึกษากระบวนการต่างๆข้างต้นมาแล้ว ขั้นตอนสุดท้ายที่ต้องพิจารณาก็คือวิธีการที่จะให้คอมพิวเตอร์ประมวลผลข้อมูลที่จัดเตรียมไว้แล้ว ให้ออกมาเป็นผลลัพธ์ที่ถูกต้องแม่นยำ วิธีการนี้เรียกว่า “แบบจำลองที่ใช้” แบบจำลองนี้ทำหน้าที่เป็นแนวทางในการประมวลผลข้อมูลจากการศึกษาที่ผ่านมามีการใช้แบบจำลองต่างๆ ในการช่วยให้คอมพิวเตอร์ปฏิบัติการได้จริง โดยใช้แบบจำลองต่างๆ ดังนี้

2.8.1 **การใช้แบบจำลองทางสถิติ** (Probabilistic model)

ในการถอดอักษรส่วนใหญ่นั้นจะใช้วิธีการทางสถิติเข้ามาช่วยโดยมีการนำวิธีทางสถิติไปประยุกต์ใช้ ต่อ เช่นใช้วิธีดังต่อไปนี้

1) Weighted finite-state transducer (WFST) หรือ เรียกว่าเป็นเครือข่ายการเปลี่ยนตำแหน่ง (Transition network) ซึ่งเมื่อมีข้อมูลเข้ามาก็จะทำให้เกิดการเปลี่ยนตำแหน่งพร้อมทั้งสร้างข้อมูลออกด้วย โดยที่การเปลี่ยนตำแหน่งจากสภาพ (state) หนึ่งไปยังอีกสภาพหนึ่งจะมีน้ำหนักแตกต่างกัน ในเรื่องของการถอดอักษรนี้ ข้อมูลเข้าอาจเป็นตัวอักษรข้อมูลออกอาจเป็นตัวอักษรในอีกภาษาหรือเป็นเสียงที่สอดคล้องกัน เป็นต้น ตัวอย่างของงานวิจัยที่ใช้วิธีนี้ เช่นในงานของKnight and Graehl (1997) ซึ่งศึกษาเรื่อง การถอดอักษรคาตาคานะในภาษาญี่ปุ่นที่เขียนทับศัพท์ภาษาอังกฤษกลับไปเป็นภาษาอังกฤษโดยใช้เครื่องคอมพิวเตอร์ ได้ใช้แนวทางการถอดอักษรโดยการใช้ Weighted finite-state transducer (WFST) มีการกำหนดน้ำหนักของคำและนำมาใช้ในการถอดอักษรให้ถูกต้องและแม่นยำ

2) **สมมติฐานของมาคอฟ** (Markov Assumption) นั้นมีการนำมาใช้ในงานการถอดอักษรมาก เนื่องจากเป็นวิธีที่ทำให้ประหยัดระยะเวลาในการปฏิบัติงานได้มาก วิธีนี้เป็นการประมาณค่าทางสถิติของข้อมูลที่ต้องการเพื่อใช้ค่าคำนวณ เช่น $P(T \mid E) \cong \prod_{i=1}^{n} P(t_i \mid e_i)$ หมายถึงค่าความน่าจะเป็นในการถอดอักษรจากคำ E เป็น T นั้น ประมาณได้จากผลคูณของความน่าจะเป็นในการแปลงจากตัวอักษร e_i เป็น t_i ตัวอย่างของงานที่ใช้วิธีนี้ เช่นงานของ Jung, Hong, and Paek (2000) ซึ่งได้ศึกษาเรื่องรูปแบบการถอดอักษรภาษาอังกฤษเป็นภาษาเกาหลีได้ใช้แนวทางการถอดอักษรโดยการเพิ่มเติมกฎของมาคอฟข้างต้นให้มีประสิทธิภาพในการใช้ถอดอักษรมากขึ้น

2.8.2 **การใช้เครือข่ายนิวรอล** (Neural Network)

การใช้เครือข่ายนิวรอลเป็นวิธีหนึ่งที่นำมาในการถอดอักษรเพื่อใช้ในการค้นคืนข้อมูลข้ามภาษา วิธีนี้มองการทำงานเป็นแบบจำลองที่ประกอบด้วยปุ่มนิวรอล (Neural nodes) จำนวนมากโยงใยกันเหมือนระบบในสมองมนุษย์ โดยรับข้อมูลเข้าเป็นรูปอักษรที่ต้องการ นำมาใส่รหัสในคำและบริบทซ้ายและขวาอย่างละ 2 ตำแหน่ง และให้ผลลัพธ์ออกมาเป็นรหัสเสียงของรูปอักษรที่รับเข้ามา ตัวอย่างงานที่ใช้วิธีนี้ เช่นงานของ ทัศนวรรณ ศูนย์กลาง, สมชาย ประสิทธิ์จู

ตระกูล, และ บุญเสริม กิจศิริกุล (2543) ซึ่งได้ศึกษาเรื่องการเข้ารหัสคำทับศัพท์ภาษาไทยและอังกฤษเพื่อค้นคืนข้ามภาษาด้วยใช้วิธีการเครือข่ายนิวรอล ได้ใช้แนวทางการนำเครือข่ายนิวรอลมาใช้ในการค้นคืนข้ามภาษาไทยและภาษาอังกฤษ

2.8.3 **การใช้แบบจำลองแบบเปรียบเทียบตารางกฎ** (Table lookup)

การใช้แบบจำลองแบบเปรียบเทียบตารางกฎนั้น คือ การนำข้อมูลที่รวบรวมเป็นคลังข้อมูลไว้นั้น มาใส่ลงในตารางที่แยกตามบริบทการเกิดของข้อมูลหรืออักษรนั้นๆ ตารางนี้จะเปรียบเสมือนกฎที่กำหนดขึ้นให้คอมพิวเตอร์เข้าใจถึงกระบวนการทำงาน เมื่อเราสร้างตารางที่แยกตามบริบทได้เรียบร้อยแล้ว คอมพิวเตอร์จะทำการเทียบข้อมูลที่รับเข้ามากับกฎที่มีในตาราง เมื่อพบอักษรที่ตรงกับตารางใดก็จะประมวลผลแต่ละตัวอักษรออกมาตามกฎที่มีอยู่ในตารางนั้น จะประมวลผลจนครบทุกตัวอักษรก็จะได้ผลลัพธ์ที่ถูกต้องตามกฎออกมา ตัวอย่างของงานที่ใช้วิธีนี้ เช่น งานของ Van den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่องการแปลงรูปอักษรเป็นรูปแทนเสียงโดยการใช้คลังข้อมูล ได้ใช้แนวทางการเทียบข้อมูลที่รับเข้ากับตารางกฎที่คอมพิวเตอร์สร้างขึ้นมา เมื่อพบรูปอักษรตรงกับข้อมูลที่รับเข้ามาก็จะแปลงเป็นรูปแทนเสียงตามกฎในตารางนั้น

ในการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้ศึกษาคำจำกัดความของการถอดอักษร (Transliteration) ประวัติความเป็นมาของการถอดอักษรในภาษาไทย หลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน และแนวทางการถอดอักษรในภาษาต่างๆโดยใช้คอมพิวเตอร์ดังรายละเอียดข้างต้น ในบทต่อไปจะขอนำเสนอเกี่ยวกับวิธีการดำเนินการวิจัย การวิเคราะห์ข้อมูลและแบบจำลองที่ใช้ในวิทยานิพนธ์ฉบับนี้โดยละเอียด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

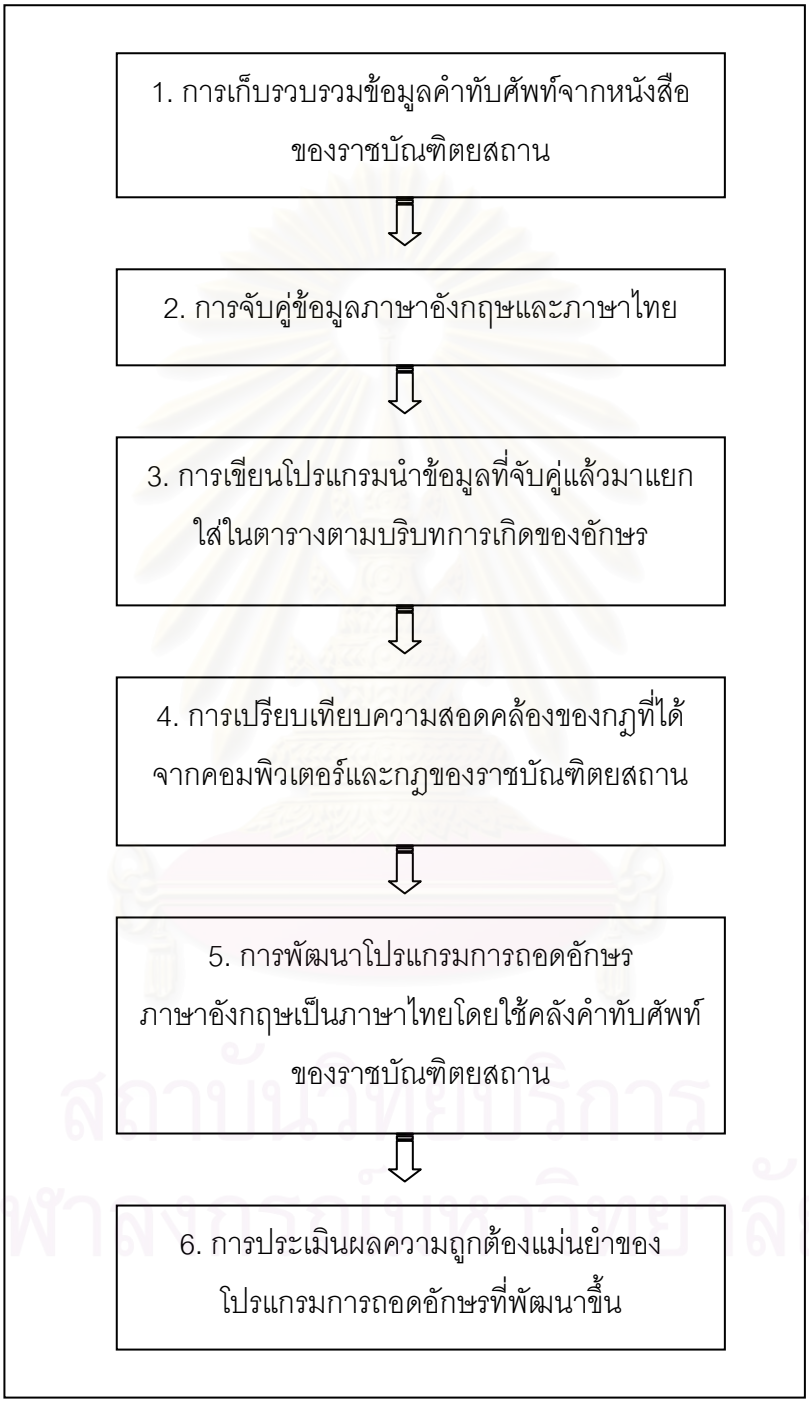
บทที่ 3

วิธีดำเนินการวิจัย

หลังจากได้ทบทวนวรรณกรรมเกี่ยวกับการวิจัยเรื่องการถอดอักษรในภาษาต่างๆ แล้ว จะเห็นได้ว่าในการเขียนโปรแกรมการถอดอักษรโดยใช้คอมพิวเตอร์นั้น จะต้องมีขั้นตอนหลายขั้นตอนเพื่อให้ได้มาซึ่งโปรแกรมการถอดอักษรที่สมบูรณ์ สำหรับวิทยานิพนธ์เล่มนี้จะเน้นความสำคัญไปที่การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน ผู้วิจัยต้องการจะใช้วิธีการถอดอักษรโดยตรง คือ ถอดอักษรจากภาษาอังกฤษเป็นภาษาไทย โดยไม่ต้องผ่านตัวกลางคือ รูปแทนเสียง เพราะเห็นว่าหลักเกณฑ์การทับศัพท์ที่ราชบัณฑิตยสถานให้ไว้นั้น ได้พิจารณาถึงความสอดคล้องทางเสียงของอักษรภาษาอังกฤษและภาษาไทยเรียบร้อยแล้ว และได้บัญญัติให้อักษรภาษาอังกฤษตัวหนึ่งถอดเป็นอักษรภาษาไทยตัวหนึ่งโดยตรง นอกจากนี้ผู้วิจัยมีความเห็นว่าคำทับศัพท์ที่เผยแพร่โดยราชบัณฑิตยสถานนั้น น่าจะเป็นคำทับศัพท์ที่ถูกต้องตามหลักเกณฑ์คำทับศัพท์ที่ราชบัณฑิตยสถานวางไว้ ดังนั้นผู้วิจัยจึงเห็นว่าหากเราเก็บคำทับศัพท์ภาษาอังกฤษ ที่มีทั้งคำภาษาอังกฤษและคำทับศัพท์ที่เขียนเป็นภาษาไทยที่พิมพ์เผยแพร่โดยราชบัณฑิตยสถานเป็นคลังข้อมูลที่มีขนาดประมาณ 10,000 คำ แล้วนำข้อมูลเหล่านี้มาจับคู่ข้อมูลและประมวลผลโดยโปรแกรมคอมพิวเตอร์ให้สร้างเป็นกฎการถอดอักษรสำหรับคอมพิวเตอร์ขึ้นมา โดยที่กฎเหล่านี้จะจำแนกลงในตารางตามแนวทางแบบจำลองแบบเปรียบเทียบตารางกฎ (Table Lookup) ก็น่าจะได้โปรแกรมการถอดอักษรที่มีความถูกต้องแม่นยำออกมา

ในส่วนของกระบวนการการพัฒนาโปรแกรมนั้นมีขั้นตอนแรกเป็นการเก็บรวบรวมข้อมูลคำทับศัพท์เพื่อสร้างเป็นคลังคำทับศัพท์ของราชบัณฑิตยสถานขนาดประมาณ 10,000 คำ ส่วนขั้นตอนต่อไปจะเป็นส่วนของการจับคู่ข้อมูลภาษาอังกฤษและภาษาไทย ถัดไปจะเป็นขั้นตอนการเขียนโปรแกรมเพื่อนำข้อมูลที่จับคู่แล้วมาใส่ลงในตารางที่เรียงตามบริบทการเกิดของตัวอักษรตามแนวทางแบบจำลองแบบเปรียบเทียบตารางกฎ (Table Lookup) จากนั้นจะนำกฎที่ได้จากคอมพิวเตอร์มาเปรียบเทียบกับกฎที่ราชบัณฑิตยสถานให้ไว้ว่ามีความสอดคล้องกันเพียงไร และต่อไปจะเป็นขั้นตอนของการพัฒนาโปรแกรมการถอดอักษรโดยการเทียบกฎจากในตารางที่มีบริบทการเกิดของอักษรเรียบร้อยแล้ว ขั้นตอนสุดท้ายจะเป็นขั้นตอนการประเมินผลความถูกต้องแม่นยำของโปรแกรมที่พัฒนาขึ้น

กระบวนการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานสามารถแสดงได้ดังรูปที่ 3-1



รูปที่ 3-1 กระบวนการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทย

โดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน

3.1 การเก็บรวบรวมข้อมูล

ในส่วนของขั้นตอนแรกในการวิจัยนี้ คือ การเก็บรวบรวมข้อมูล ข้อมูลที่ต้องการก็คือคำทับศัพท์ภาษาอังกฤษเป็นภาษาไทยที่พิมพ์เผยแพร่โดยราชบัณฑิตยสถาน โดยข้อมูลที่เก็บนี้จะเก็บเป็นคู่คำภาษาอังกฤษและคำทับศัพท์ที่เขียนเป็นภาษาไทยจำนวนประมาณ 10,000 คู่คำ ซึ่งรวบรวมมาจากหนังสือของราชบัณฑิตยสถานดังต่อไปนี้

1) ความรู้ทางอักษรศาสตร์
2) พจนานุกรมศัพท์ยานยนต์
3) พจนานุกรมศัพท์ธรณีวิทยา
4) พจนานุกรมศัพท์ภูมิศาสตร์ อังกฤษ-ไทย ฉบับราชบัณฑิตยสถาน เล่ม ๑ A-K
5) พจนานุกรมศัพท์ภูมิศาสตร์ อังกฤษ-ไทย ฉบับราชบัณฑิตยสถาน เล่ม ๒ L-Z
6) พจนานุกรมศัพท์ปรับอากาศ
7) พจนานุกรมศัพท์ยานยนต์
8) พจนานุกรมศัพท์วรรณกรรม อังกฤษ-ไทย
9) พจนานุกรมศัพท์สัทศาสตร์
10) สารานุกรมประวัติศาสตร์สากล: อเมริกา เล่ม ๑ อักษร A-B
11) สารานุกรมประวัติศาสตร์สากลสมัยใหม่: ยุโรป เล่ม ๒ อักษร C-D
12) สารานุกรมประวัติศาสตร์สากลสมัยใหม่: เอเชีย เล่ม ๑ อักษร A-B
13) ศัพท์การเชื่อม
14) ศัพท์คณิตศาสตร์
15) ศัพท์เทคโนโลยีสารสนเทศฉบับราชบัณฑิตยสถาน
16) ศัพท์พลังงาน อังกฤษ-ไทย
17) ศัพท์ภาษาศาสตร์
18) ศัพท์วิทยาศาสตร์
19) หลักเกณฑ์การทับศัพท์ภาษาอังกฤษ

ข้อมูลคำทับศัพท์ที่เก็บเข้ามานั้น ในขั้นแรกจะเก็บข้อมูลลงในโปรแกรม Microsoft Excel โดยที่ในสดมภ์แรกจะเก็บเป็นคำภาษาอังกฤษ และในสดมภ์ที่ 2 จะเก็บเป็นคำภาษาไทย ผู้วิจัยจะเก็บข้อมูลคำทับศัพท์จากหนังสือแต่ละเล่มของราชบัณฑิตยสถานไว้ในแผ่นงานที่แยกจากกัน เพื่อ

ในกรณีที่ต้องการกลับมาตรวจสอบความถูกต้องของข้อมูลจะได้รู้ว่าคำที่สงสัยนั้นมาจากหนังสือเล่มใด ข้อมูลที่เก็บมานั้นแสดงเป็นตัวอย่างได้ดังนี้

abercrombie	แอเบอร์ครอมบี
arytenoid	แอริทินอยด์
cricoid	ไครคอยด์
Daniel	แดเนียล
david	เดวิด
electromyograph	อิเล็กทรอไมโอกราฟ
Haas	ฮาส
harmonic	ฮาร์มอนิก
Jones	โจนส์
Ladefoged	แลเดโฟเกด

ตารางที่ 3-1 ตัวอย่างข้อมูลที่เก็บมาจากพจนานุกรมศัพท์สัทศาสตร์

acute	อะคิวต์
Alpha	อัลฟา
Indian	อินเดียน
Aramaic	อาราเมอิก
argument	อาร์กิวเมนต์
Aryan	อารยัน
arytenoid	แอริทินอยด์
Australian	ออสเตรเลียน
Austric	ออสตริก
Austroasiatic	ออสโตรเอเชียติก

ตารางที่ 3-2 ตัวอย่างข้อมูลที่เก็บมาจากศัพท์ภาษาศาสตร์

ข้อมูลที่รวบรวมมาทั้งหมดจำนวนประมาณ 10,000 คำ จะจัดเก็บอยู่ในแผ่นงาน (Worksheet) ทั้งหมด 19 แผ่นงาน โดยจำแนกตามรายชื่อหนังสือ หลังจากนั้นจะนำข้อมูลทั้งหมดที่เก็บได้มา

รวมกันในแผ่นงานเดียวกันโดยเรียงลำดับจากตัวอักษร A ไปถึงอักษร Z แล้วจึงทำการแปลงชนิดของไฟล์จากชนิดสมุดงาน Microsoft Excel เป็นชนิดข้อความที่มีแท็บเป็นตัวคั่น เพื่อที่จะนำข้อมูลนี้ไปประมวลผลโดยโปรแกรมภาษา Perl ในขั้นตอนต่อไป

3.2 การจับคู่ข้อมูลภาษาอังกฤษและภาษาไทย

หลังจากที่ได้ข้อมูลที่มีลักษณะเป็นคู่คำภาษาอังกฤษและภาษาไทยจำนวนประมาณ 10,000 คำที่จัดเก็บเป็นชนิดข้อความโดยใช้แท็บเป็นตัวคั่นเรียบร้อยแล้ว ก็จะนำข้อมูลที่ได้นั้นมาจับคู่ข้อมูลภาษาอังกฤษและภาษาไทย ด้วยวิธีการจับคู่ข้อมูลแบบอัตโนมัติ (Automatic alignment) นั่นคือ การนำคำภาษาอังกฤษและภาษาไทยที่เก็บรวบรวมมานั้น มาจับคู่อักษรอังกฤษและไทย โดยใช้โปรแกรมการจับคู่ข้อมูลภาษาอังกฤษและภาษาไทย เหตุผลที่ใช้วิธีจับคู่ข้อมูลแบบอัตโนมัตินั้นก็เพราะว่าเป็นวิธีที่สะดวกและรวดเร็ว

สำหรับโปรแกรมการจับคู่ข้อมูลนี้จะนำคำภาษาอังกฤษและภาษาไทยที่เก็บมานั้นมาใส่ตัวคั่นอักษร โดยที่อักษรอังกฤษและภาษาไทยที่ใส่ตัวคั่นแล้วในตำแหน่งเดียวกันจะต้องสอดคล้องกันทางเสียง ผู้วิจัยกำหนดให้โปรแกรมตัดอักษรภาษาอังกฤษและภาษาไทยโดยใช้เครื่องหมาย “/” เป็นตัวคั่น ในส่วนของข้อมูลภาษาอังกฤษนั้นกำหนดให้โปรแกรมตัดอักษรให้ใกล้เคียงตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถานเพื่อที่จะทำให้สามารถมาเปรียบเทียบกับกฎที่ราชบัณฑิตยสถานวางไว้ได้สะดวกในภายหลัง การตัดอักษรจะใส่ตัวคั่น “/” หลังอักษรต่อไปนี้

สำหรับพยัญชนะภาษาอังกฤษจะใส่ตัวคั่นเมื่อพบอักษรต่อไปนี้
thm,sch,ch,ck,gh,gn,kh,ph,rh,sh,th,wh,b,c,d,f,g,h,j,k,l,m,n,p,q,r,s,t,v,w,x,y,z

สำหรับสระภาษาอังกฤษจะใส่ตัวคั่นเมื่อพบอักษรต่อไปนี้
ai,ee,ea,eau,ei,eu,ie,io,oa,oe,oo,ou,ui,aea,ae,eou,aa,ao,au,eo,eu,ia,iu,oi,ua,ue,a,e,i,o,u,y

ในส่วนของภาษาไทยนั้น สำหรับพยัญชนะภาษาไทยจะใส่ตัวคั่นกับพยัญชนะทุกตัวคือ
ก,ข,ค,ง,จ,ฉ,ช,ซ,ฌ,ฎ,ฏ,ฐ,ฑ,ฒ,ณ,ด,ต,ถ,ท,ธ,น,บ,ป,ผ,ฝ,พ,ฟ,ภ,ม,ย,ร,ล,ว,ศ,ษ,ส,ห,ฬ,อ,ฮ

สำหรับสระภาษาไทยนั้นหากเป็นสระเดี่ยวที่ตามหลังพยัญชนะต้นหรือสระเดี่ยวที่อยู่บนหรือล่างพยัญชนะต้น เช่น สระอะ, สระอา, สระอิ, สระอี, สระอุ หรือสระอู จะใส่ตัวคั่นด้านหลังได้เลยแต่สำหรับสระที่เหลือจะพบปัญหาในการตัดอักษรดังต่อไปนี้

1) **สระหน้า เ,แ,โ,ไ,ใ** พบปัญหาคือเมื่อต้องการจับคู่ข้อมูลกับภาษาอังกฤษแล้วตำแหน่งของสระภาษาอังกฤษและภาษาไทยจะไม่ตรงกัน สำหรับภาษาอังกฤษนั้น สระจะตามหลังพยัญชนะต้นเสมอ แต่ภาษาไทยนั้นมีทั้งสระเดี่ยวที่อยู่หน้าหรือหลังพยัญชนะต้น จึงจำเป็นต้องย้ายสระที่อยู่ข้างหน้าพยัญชนะไปไว้ข้างหลังพยัญชนะเพื่อที่จะสามารถจับคู่กับสระภาษาอังกฤษได้ สระหน้าที่ต้องย้ายไปหลังพยัญชนะได้แก่ เ,แ,โ,ไ,ใ โดยที่เมื่อย้ายไปไว้ข้างหลังแล้วจะใส่จุดให้ 1 จุดหากพยัญชนะต้นมีอักษรเดียว หากพยัญชนะต้นเป็นคำควบกล้ำจะใส่จุด 2 จุด เพื่อในกรณีที่จะนำไปใช้ในการถอดอักษรจะได้นำพยัญชนะข้างหน้าสระมาใส่แทนที่เครื่องหมาย “.” ได้ถูกต้อง เช่น คำว่า “byte” ที่ถอดอักษรเป็น “ไบต์” เมื่อจับคู่ข้อมูลแล้วจะเป็น b/y/t/e/ และ บ/ไ./ต์/-/

2) **สระประสม เ-ะ,เ็-,แ-ะ,แ็-,เ-อ,เ-ิ,เ-อะ,โ-ะ,เ-าะ,เ-ีย, ,เ-ือ,เ-า, -ำ** พบปัญหาคือเมื่อต้องการนำสระประสมภาษาไทยมาจับคู่กับสระประสมภาษาอังกฤษ จะเห็นได้ว่าสระประสมภาษาไทยนั้นอาจจะมีทั้งอยู่หน้าและหลังพยัญชนะต้นหรืออาจจะมีที่อยู่บนพยัญชนะต้นด้วย ฉะนั้นเมื่อต้องการนำมาจับคู่กับสระประสมภาษาอังกฤษแล้วจึงต้องรวบสระประสมภาษาไทยให้อยู่ในตำแหน่งเดียวกัน ดังนั้นสระประสมข้างต้นจะถูกรวมเป็น 1 ตำแหน่งโดยที่พยัญชนะต้นจะถูกแทนที่ด้วยเครื่องหมาย “.” หากมีพยัญชนะต้นตัวเดียวก็จะมี 1 จุด หากเป็นพยัญชนะควบกล้ำก็จะมี 2 จุดตามลำดับ เช่น คำว่า “cadmium” ที่ถอดอักษรเป็น “แคดเมียม” เมื่อจับคู่ข้อมูลแล้วจะเป็น c/a/d/m/iu/m/ และ ค/แ./ด/ม/เ.ีย/ม/

3) **สระที่มีพยัญชนะต้นเป็น ‘อ’** พบปัญหาคือเมื่อต้องการจับคู่กับอักษรภาษาอังกฤษแล้ว ในส่วนของภาษาอังกฤษจะใช้สระขึ้นต้นคำได้เลย แต่ในส่วนของภาษาไทยจะมี ‘อ’ เป็นพยัญชนะต้น จึงต้องทำการรวม ‘อ’ กับสระ เพื่อที่จะได้จับคู่ในตำแหน่งเดียวกัน เช่น คำว่า “Ababa” ที่ถอดอักษรเป็น “อาบาบา” เมื่อจับคู่ข้อมูลแล้วจะเป็น a/b/a/b/a/และ อา/บ/า/บ/า/

เมื่อทดลองใช้โปรแกรมจับคู่อักษรแบบอัตโนมัติแล้วพบว่ายังมีข้อผิดพลาดในการตัดอักษรอยู่พอสมควร เช่น มีอักษรภาษาอังกฤษหลายตัวไม่ออกเสียง เช่น e,ue และ gh เป็นต้น ทำให้จับคู่กับภาษาไทยได้ไม่พอดีตำแหน่ง จึงต้องใส่เครื่องหมาย “-“ สำหรับอักษรภาษาอังกฤษที่ไม่ออกเสียงและไม่ปรากฏรูปในภาษาไทย นอกจากนี้อักษรภาษาอังกฤษบางตัว ถอดเป็นอักษรไทยซ้ำกัน 2 ตัว เช่น ในคำว่า double ถอดได้เป็น ดับเบิล ‘b’ ตัวเดียวจะตรงกับ ‘บ’ 2 ตัว อีกทั้งคำที่ลงท้ายด้วย le และ re เวลาถอดอักษรแล้วจะสลับกับภาษาไทย เช่น ‘double’ เป็น ‘ดับเบิล’ นั้น ‘l’ จะไปจับคู่ตรงกับ ‘เ. ‘ และ ‘e’ จะจับคู่ตรงกับ ‘ล’ และ ‘acre’ เป็น ‘เอเคอร์’ นั้น ‘r’ จะจับคู่ตรงกับ ‘ เ-อ’ และ ‘e’ จะจับคู่ตรงกับ ‘ร์’ ซึ่งเป็นการจับคู่ที่ไม่ถูกต้อง ดังนั้นผู้วิจัยจึงต้องตรวจสอบและแก้ไขการจับคู่ข้อมูลด้วยมือ (Manual Alignment) อีกครั้งเป็นขั้นตอนสุดท้าย เพื่อให้ได้ผลการจับคู่ข้อมูลที่ถูกต้องและแม่นยำขึ้น ข้อมูลที่จับคู่เรียบร้อยแล้วจะมีลักษณะดังเช่นตาราง 3-3

a/c/r/e/	เอ/ค/ร์/เ.อ./
a/m/i/s/	เอ/มิ//ส/
a/m/m/e/t/e/r/	แอ/ม/มิ//ต/เ.อ/ร์/
a/m/m/o/n/i/t/e/	แอ/ม/ม/โ./น/ไ./ต์/-/
a/m/m/o/n/ia/	แอ/ม/ม/โ./น/เีย/
a/m/m/o/n/oi/d/	แอ/ม/ม/โ./น/อย/ด์/
a/m/n/i/o/t/e/	แอ/ม/นิ/โอ/ต/-/
a/m/o/r/g/o/s/	อะ/ม/อ/ร์/ก/อ/ส/
a/m/o/s/	เอ/ม/อ/ส/
c/i/r/q/u/e/	ซ/เิ./ร์/ก/-/-/
d/ou/b/l/e/	ดั/บบ/ล/เ../
h/i/gh/	ฮ/ไ./-/

ตาราง 3-3 ตัวอย่างข้อมูลที่จับคู่ข้อมูลแล้ว

จากตาราง 3-3 จะเห็นได้ว่าข้อมูลภาษาอังกฤษและภาษาไทยที่จับคู่ข้อมูลเรียบร้อยแล้วจะมีลักษณะที่ตำแหน่งของอักษรภาษาอังกฤษและภาษาไทยที่มีความสอดคล้องกันทางเสียงจะอยู่ตำแหน่งเดียวกัน เช่น คำว่า amis นั้น ‘a’ จะตำแหน่งเดียวกับ ‘เอ’ ส่วน ‘m’จะตำแหน่งเดียวกับ ‘ม’ สำหรับ ‘i’จะตำแหน่งเดียวกับ ‘ ิ’ และ ‘s’จะตำแหน่งเดียวกับ ‘ส’

เมื่อนำคู่คำประมาณ 10,000 คู่คำมาจับคู่ข้อมูลเรียบร้อยแล้ว จะพบว่ามีคำภาษาอังกฤษจำนวนหลายคำที่สามารถถอดอักษรเป็นภาษาไทยได้หลายแบบ เช่น คำว่า gas ที่ถอดอักษรได้ทั้งเป็นคำว่า “ก๊าซ” หรือ “แก๊ส” ผู้วิจัยเห็นว่าหากเก็บการถอดอักษรหลายแบบนี้เข้าไปในคลังข้อมูลจะส่งผลกระทบต่อผลการถอดอักษรให้ไม่แม่นยำ เนื่องจากผู้วิจัยต้องการให้เมื่อใส่คำภาษาอังกฤษเข้าไปแล้ว จะได้การถอดอักษรภาษาไทยที่ถูกต้องออกมาแบบเดียว จึงทำการคัดเลือกแบบการถอดอักษรที่ได้มาตรฐานที่สุดเพียงแบบเดียว โดยเทียบจากหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน และเทียบปี พ.ศ. ที่ข้อมูลนั้นพิมพ์เผยแพร่ออกมา หากพบว่าข้อมูลนั้นมาจากหนังสือของราชบัณฑิตยสถานที่พิมพ์เผยแพร่ก่อนปี พ.ศ. 2532 ซึ่งเป็นปีที่ราชบัณฑิตยสถานได้พิมพ์เผยแพร่หนังสือหลักเกณฑ์การทับศัพท์ภาษาอังกฤษ ก็จะตัดข้อมูลนั้นทิ้งไป นอกจากนี้ข้อมูลที่พิมพ์เผยแพร่ก่อนปี พ.ศ. 2532 ทั้งหมดก็จะถูกตัดทิ้งไปเพราะพบว่าการถอดอักษรของหนังสือก่อนปี พ.ศ. 2532 นั้นส่วนใหญ่ไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ภาษาอังกฤษฉบับราชบัณฑิตยสถานที่พิมพ์เผยแพร่ในปี พ.ศ. 2532 สรุปแล้วจากข้อมูลทั้งหมด 10,060 คู่คำ ถูกตัดออกไปทั้งสิ้น 1,870 คู่คำ คงเหลือข้อมูล 8,190 คู่คำในคลังข้อมูล หลังจากได้จับคู่ข้อมูลและเตรียมข้อมูลเรียบร้อยแล้วก็จะนำข้อมูลเหล่านี้ไปแยกใส่ตารางในขั้นตอนต่อไป

3.3 การเขียนโปรแกรมนำข้อมูลใส่ตารางตามบริบท

หลังจากที่ได้ข้อมูลที่มีจากจับคู่ข้อมูลภาษาอังกฤษและภาษาไทยให้อักษรที่มีความสอดคล้องกันทางเสียงจับคู่ในตำแหน่งเดียวกันแล้ว สิ่งที่จะทำเป็นขั้นตอนต่อไปก็คือ การพัฒนาโปรแกรมเพื่อประมวลผลข้อมูลที่จับคู่แล้วมาจำแนกลงตารางที่เรียงลำดับตามบริบทการถอดอักษรภาษาอังกฤษเป็นภาษาไทยตามแนวทาง แบบจำลองแบบเปรียบเทียบตารางกฎ (Table Lookup) ในงานของVan den Bosch and Daelemans (1993) ซึ่งได้ศึกษาเรื่องการแปลงรูปอักษรเป็นรูปแทนเสียงโดยการใช้คลังข้อมูล ได้ใช้แนวทางการเทียบข้อมูลที่รับเข้ากับตารางกฎที่คอมพิวเตอร์สร้างขึ้นมา เมื่อพบรูปอักษรตรงกับข้อมูลที่รับเข้ามาก็จะแปลงเป็นรูปแทนเสียงตามกฎในตารางนั้น บริบทการถอดอักษรนี้ก็คือ ตำแหน่งซ้ายหรือขวาของตัวอักษรที่มีอิทธิพลทำให้อักษรภาษาอังกฤษตัวหนึ่ง ถอดอักษรได้เป็นอักษรภาษาไทยตัวหนึ่ง สำหรับวิทยานิพนธ์เล่มนี้ผู้วิจัยจะนำข้อมูลที่มีการจับคู่ข้อมูลเรียบร้อยแล้ว มาสร้างเป็นตารางหลักๆ 3 ตาราง คือ

1. ตารางที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่งในการถอดอักษรเป็นภาษาไทย มีจำนวนกฎ 2, 743 กฎ ตัวอย่างเช่น

บริบทซ้าย	อักษรเป้าหมาย	บริบทขวา	อักษรที่ถอดได้
y	**c**	n	**ก**
k	**r**	i	**ร**
-	**i**	l	**อิ**
s	**w**	a	**ว**
f	**n**	i	**น**

ตาราง 3-4ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย1 ตำแหน่งและขวา 1 ตำแหน่ง

2. ตารางที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่งในการถอดอักษรเป็นภาษาไทย มีจำนวนกฎ 12,620 กฎ ตัวอย่างเช่น

บริบทซ้าย ตำแน่งที่ 2	บริบทซ้าย ตำแน่งที่ 1	อักษร เป้าหมาย	บริบทขวา ตำแน่งที่ 1	บริบทขวา ตำแน่งที่ 2	อักษรที่ถอดได้
n	a	**t**	i	t	**ท**
h	o	**r**	a	c	**ร**
c	i	**u**	s	-	**อั**
n	i	**ck**	k	e	**ก**
e	a	**e**	e	-	-

ตาราง 3-5ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3) ตารางที่อักษรภาษาอังกฤษต้องใช้บริบทซ้าย 3 ตำแหน่งและขวา 3 ตำแหน่งในการถอดอักษรเป็นภาษาไทย มีจำนวนกฎ 3,879 กฎ ตัวอย่างเช่น

บริบทซ้ายตำแน่งที่ 3	บริบทซ้ายตำแน่งที่ 2	บริบทซ้ายตำแน่งที่ 1	อักษรเป้าหมาย	บริบทขวาตำแน่งที่ 1	บริบทขวาตำแน่งที่ 2	บริบทขวาตำแน่งที่ 3	อักษรที่ถอดได้
-	-	-	p	o	r	a	พ
-	-	m	a	s	e	f	เ.
e	t	a	ph	h	a	s	ฟ
-	-	m	o	d	e	m	โ.
a	l	i	s	m	-	-	ซึ

ตาราง 3-6 ตัวอย่างข้อมูลในตารางที่มีบริบทซ้าย 3 ตำแหน่งและขวา 3 ตำแหน่ง

นอกจากตารางกฎ 3 ตารางที่สร้างขึ้นมาจากการจับคู่อักษรที่สอดคล้องกันในลักษณะ 1 อักษรต่อ 1 อักษรแล้ว ผู้วิจัยพบว่ามีอักษรเป้าหมายหลายตัวที่ในบริบทเดียวกันนั้นสามารถถอดอักษรได้เป็นหลายแบบ ซึ่งจะไม่สามารถสรุปเป็นกฎลงในตารางทั้งสามนี้ได้ จึงคัดเลือกแบบที่พบถี่ที่สุดมาเก็บไว้ในตารางคาดเดาตัวอักษรที่จะถอด ในที่นี้จะแบ่งตารางคาดเดาเป็น 3 ตาราง ดังนี้

1. ตารางคาดเดาโดยไม่ต้องใช้บริบทในการถอดอักษร ด้วยวิธีคัดเลือกแบบที่อักษรเป้าหมายภาษาอังกฤษตัวหนึ่งจะถอดอักษรเป็นภาษาไทยได้โดยที่ไม่อาศัยบริบทซ้ายขวาใดๆที่พบถี่ที่สุดในคลังข้อมูล มีจำนวนกฎ 82 กฎ ตัวอย่างเช่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อักษรเป้าหมาย	อักษรที่ถอดได้
b	บ
d	ด
f	ฟ

g	ก
h	ฮ

ตาราง 3-7 ตัวอย่างข้อมูลในตารางที่ใช่เดาโดยไม่อาศัยบริบท

2. ตารางคาดเดาที่ใช้บริบทซ้าย 1 ตำแหน่ง บริบทขวา 1 ตำแหน่ง โดยคัดเลือกจากแบบการถอดอักษรที่มีบริบทดังกล่าวที่พบถี่ที่สุดในคลังข้อมูล มีจำนวนกฎ 3,362 กฎ ตัวอย่างเช่น

บริบทซ้าย	อักษรเป้าหมาย	บริบทขวา	อักษรที่ถอดได้
m	a	y	เ
m	a	z	ะ
l	y	y	ย์
u	f	f	ฟ
g	u	a	ว

ตาราง 3-8 ตัวอย่างข้อมูลในตารางคาดเดาที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง

3. ตารางคาดเดาที่ใช้บริบทซ้าย 2 ตำแหน่ง บริบทขวา 2 ตำแหน่ง โดยคัดเลือกจากแบบการถอดอักษรที่มีบริบทดังกล่าวที่พบถี่ที่สุดในคลังข้อมูล มีจำนวนกฎ 9,019 กฎ ตัวอย่างเช่น

บริบทซ้าย ตำแน่งที่ 2	บริบทซ้าย ตำแน่งที่ 1	อักษร เป้าหมาย	บริบทขวา ตำแน่งที่ 1	บริบทขวา ตำแน่งที่ 2	อักษรที่ ถอดได้
-	-	l	e	w	ล
-	e	n	n	s	น์
i	n	sh	h	e	ช
-	s	t	u	a	จ
r	b	o	n	-	อ

ตาราง 3-9 ตัวอย่างข้อมูลในตารางคาดเดาที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง

จะเห็นได้ว่าตาราง 6 ตารางข้างต้นนั้น เปรียบเสมือนกฎให้กับคอมพิวเตอร์ใช้ คือ การกำหนดกฎการถอดอักษรให้คอมพิวเตอร์โดยใช้ฐานข้อมูลของคำที่ใช้จริง เมื่อได้กฎดังกล่าวมาแล้วจะนำมาเปรียบเทียบความสอดคล้องกับกฎที่ราชบัณฑิตยสถานกำหนดไว้ต่อไป

3.4 การเปรียบเทียบความสอดคล้องของกฎการถอดอักษรที่ได้จากคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน

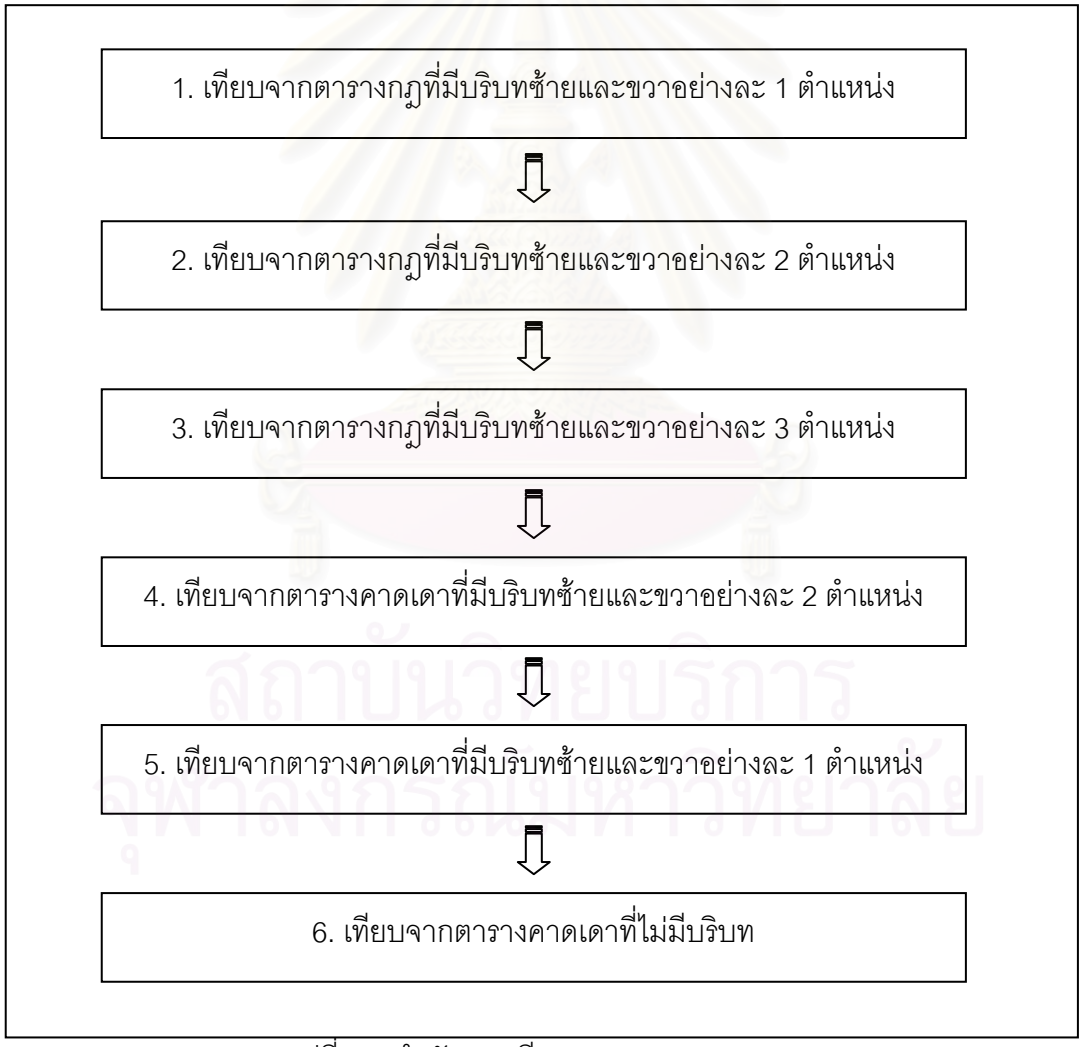
หลังจากที่ได้ตารางกฎการถอดอักษรสำหรับคอมพิวเตอร์ดังที่ได้กล่าวไปแล้วนั้น ในขั้นตอนต่อไป ผู้วิจัยจะนำกฎของคอมพิวเตอร์ที่ได้นั้นมาเปรียบเทียบความสอดคล้องกับหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน ความสอดคล้องนี้จะพิจารณาว่ากฎการถอดอักษรของคอมพิวเตอร์ที่สร้างมาจากคลังคำทับศัพท์ของราชบัณฑิตยสถานนั้นถูกต้องตรงกับหลักเกณฑ์การทับศัพท์ที่ราชบัณฑิตยสถานได้วางไว้เพียงใด โดยจะเปรียบเทียบให้เห็นเป็นจำนวนร้อยละว่าจากกฎที่ราชบัณฑิตยสถานให้ไว้นั้น จากข้อมูลจริงที่พบนั้นเป็นไปตามกฎนั้นร้อยละเท่าใด หากเป็นไปตามสมมติฐานแล้วนั้น กฎที่ได้จากในตารางกฎของคอมพิวเตอร์และกฎที่ราชบัณฑิตยสถานให้ไว้จะต้องมีความสอดคล้องกันเพราะกฎในตารางที่ได้มานั้นก็นำมาจากข้อมูลคำทับศัพท์ซึ่งราชบัณฑิตยสถานเป็นผู้เขียนขึ้นเอง

เมื่อได้เปรียบเทียบความสอดคล้องเรียบร้อยแล้ว ขั้นตอนสุดท้ายที่จะทำก็คือนำตารางกฎการถอดอักษรที่ได้จากคอมพิวเตอร์นั้นไปใช้ในการพัฒนาโปรแกรมการถอดอักษรที่สมบูรณ์ต่อไป

3.5 การพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน

เมื่อผ่านกระบวนการทั้งหมดข้างต้น มาจนถึงขั้นตอนสุดท้ายนี้แล้ว สิ่งที่ต้องทำในขั้นตอนนี้ก็คือการเขียนโปรแกรมที่เทียบคำภาษาอังกฤษที่รับมากับตารางกฎดังกล่าวข้างต้น โดยจะเทียบอักษรแต่ละตัวในคำที่รับมาโดยเริ่มจากตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 1 ตำแหน่ง (ตาราง 3-4) ก่อนหากพบข้อมูลในตารางแรกนี้ก็จะไม่จำเป็นต้องไปดูในตารางต่อไปเพราะได้ผล

ลัพธ์เรียบร้อยแล้ว แต่หากไม่พบจะไปดูตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 2 ตำแหน่ง (ตาราง 3-5) หากยังไม่พบจะไปดูตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 3 ตำแหน่ง (ตาราง 3-6) ถ้าไม่พบอักษรเป้าหมายในตารางกฎทั้งสามข้างต้นนี้จะใช้การคาดเดาจากตารางคาดเดาที่มีบริบทซ้ายและขวาอย่างละ 2 ตำแหน่ง (ตาราง 3-9) หากยังไม่พบจะไปดูตารางคาดเดาที่มีบริบทซ้ายและขวาอย่างละ 1 ตำแหน่ง (ตาราง 3-8) สุดท้ายหากเทียบอักษรจากทุกตารางข้างต้นไม่พบจะไปดูตารางคาดเดาที่ไม่มีบริบทที่นับอักษรที่มีความถี่สูงที่สุดในคลังข้อมูล (ตาราง 3-7) จึงทำให้อักษรภาษาอังกฤษทุกตัวที่รับเข้ามาสามารถถอดอักษรเป็นภาษาไทยได้ครบทุกตัวอักษร ดังแสดงให้เห็นได้ดังรูป 3-2



รูปที่ 3-2 ลำดับการเทียบตารางกฎและตารางคาดเดา

จะเห็นได้ว่าขั้นตอนการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานนั้นนอกจากจะอาศัยคลังคำทับศัพท์ของราชบัณฑิตยสถานเพื่อนำมาสร้างเป็นตารางกฎให้กับคอมพิวเตอร์ตามแนวทาง แบบจำลองแบบเปรียบเทียบตารางกฎ (Table Lookup) แล้ว ยังต้องอาศัยการนับความถี่ของการถอดอักษรช่วยในการทำให้การถอดอักษรภาษาอังกฤษเป็นภาษาไทยนั้นถูกต้องแม่นยำมากขึ้นด้วย

หลังจากได้โปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยเรียบร้อยแล้ว ก็ต้องมีการทดสอบว่าโปรแกรมที่พัฒนาขึ้นมานั้นปฏิบัติการได้ดี มีประสิทธิภาพเพียงไร จึงจะขอกล่าวถึงการประเมินผลความถูกต้องแม่นยำของโปรแกรมการถอดอักษรที่พัฒนาขึ้นมานี้ในหัวข้อถัดไป

3.6 การประเมินผลความถูกต้องแม่นยำของโปรแกรมการถอดอักษร

การที่จะทดสอบว่าโปรแกรมการถอดอักษรที่พัฒนาขึ้นมานี้ทำงานได้ดีเพียงใดนั้น ก็ต้องมีการเปรียบเทียบความถูกต้องของคำที่ถอดอักษรได้จากโปรแกรมนี้กับคำที่ถอดอักษรโดยราชบัณฑิตยสถาน สำหรับวิทยานิพนธ์เล่มนี้ ผู้วิจัยได้ทำการแบ่งข้อมูลทั้งหมดจำนวนประมาณ 10,000 คู่คำ ออกเป็น 2 ส่วนโดยการสุ่มเลือก (random) ดังนี้

1) ส่วนของคลังข้อมูลฝึกสอนซึ่งมีจำนวนคำร้อยละ 80 ของข้อมูลทั้งหมด

2) คลังข้อมูลทดสอบซึ่งมีจำนวนคำร้อยละ 20 ของข้อมูลทั้งหมด

หลังจากแบ่งข้อมูลเรียบร้อยแล้ว ก็นำตารางกฎที่ได้จากคลังข้อมูลฝึกสอนมาทดสอบการถอดอักษรกับคลังข้อมูลทดสอบ แล้วดูว่าคำที่ถอดอักษรออกมานั้นเหมือนกับที่ราชบัณฑิตยสถานพิมพ์เผยแพร่ไว้หรือไม่ โดยเทียบผลความถูกต้องเป็นร้อยละ โดยจะวัดผลในสองแบบ แบบแรกคือดูทั้งคำว่าตรงกับคำทับศัพท์ที่ราชบัณฑิตยสถานเขียนหรือไม่ แบบที่สองคือดูทีละตัวอักษรว่าตัวอักษรอังกฤษนั้นถอดเป็นอักษรไทยได้ถูกต้องหรือไม่

นอกจากนี้เพื่อป้องกันการเกิดความลำเอียงจากคลังข้อมูลฝึกสอน จึงต้องทำการทดสอบซ้ำแต่เปลี่ยนคลังข้อมูลฝึกสอน ผู้วิจัยจึงนำข้อมูลทั้งหมดมาสุ่มเลือกแบ่งเป็น 5 ชุด โดยแต่ละชุดจะมีจำนวนคำประมาณ 1,638 คำ และทำการทดสอบทั้งหมด 5 ครั้ง ดังนี้

1) การทดสอบครั้งที่ 1 ทำโดยการนำข้อมูลชุดที่ 1,2,3 และ 4 มาเป็นคลังข้อมูลฝึกสอนและใช้ข้อมูลชุดที่ 5 เป็นคลังข้อมูลทดสอบ

2) การทดสอบครั้งที่ 2 ทำโดยการนำข้อมูลชุดที่ 2,3,4 และ 5 มาเป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 1 เป็นคลังข้อมูลทดสอบ

3) การทดสอบครั้งที่ 3 ทำโดยการนำข้อมูลชุดที่ 3,4,5 และ 1 มาเป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 2 เป็นคลังข้อมูลทดสอบ

4) การทดสอบครั้งที่ 4 ทำโดยการนำข้อมูลชุดที่ 4,5,1 และ 2 มาเป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 3 เป็นคลังข้อมูลทดสอบ

5) การทดสอบครั้งที่ 5 ทำโดยการนำข้อมูลชุดที่ 5,1,2 และ 3 มาเป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 4 เป็นคลังข้อมูลทดสอบ

เมื่อทำการทดสอบครบ 5 ครั้งแล้วจะนำผลของความถูกต้องแม่นยำในแต่ละครั้งมาเฉลี่ยหาค่าความถูกต้องของโปรแกรมการถอดอักษรที่พัฒนาขึ้นมานี้

ขั้นตอนทั้งหมดที่ได้กล่าวไปเป็นวิธีการดำเนินการวิจัยโดยละเอียด เมื่อได้ทำขั้นตอนดังกล่าวหมดแล้ว ก็จะได้โปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน ผลการเปรียบเทียบความสอดคล้องของกฎที่ได้จากคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน และผลความถูกต้องแม่นยำของโปรแกรมที่พัฒนาขึ้นมา ในบทต่อไปจะกล่าวถึงผลการถอดอักษรโดยใช้โปรแกรมคอมพิวเตอร์ที่ผู้วิจัยพัฒนาขึ้น ซึ่งประกอบด้วยผลการเปรียบเทียบความสอดคล้องของกฎที่ได้จากคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน ผลการถอดอักษรโดยโปรแกรมการถอดอักษรและผลความถูกต้องแม่นยำของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการถอดอักษรโดยใช้คอมพิวเตอร์

ในบทที่ผ่านมาได้กล่าวถึงประเด็นทั้งหมดที่เกี่ยวข้องกับการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานไปแล้ว ในบทนี้จะกล่าวถึงการนำโปรแกรมที่พัฒนาขึ้นมาทดลองถอดอักษรภาษาอังกฤษเป็นภาษาไทยและเปรียบเทียบความสอดคล้องของกฎการถอดอักษรของคอมพิวเตอร์กับกฎการถอดอักษรของราชบัณฑิตยสถาน อีกทั้งกล่าวถึงผลลัพธ์จากการถอดอักษรด้วยโปรแกรมและผลความถูกต้องแม่นยำของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมา

4.1 ผลการเปรียบเทียบความสอดคล้องของกฎที่ได้จากคอมพิวเตอร์และกฎของราชบัณฑิตยสถาน

จากบทที่ 3 จะเห็นได้ว่าขั้นตอนในการสร้างกฎให้กับคอมพิวเตอร์นั้น ทำได้โดยการนำข้อมูลที่มีการจับคู่อักษรไว้แล้วไปประมวลผลโดยโปรแกรมและจะสร้างเป็นตารางที่เรียงลำดับตามบริบทซ้ายขวาออกมา ตารางเหล่านี้เองที่เป็นเหมือนกฎให้คอมพิวเตอร์เข้าใจว่าจะถอดอักษรภาษาอังกฤษตัวหนึ่งเป็นอักษรภาษาไทยอีกตัวหนึ่งในบริบทใด กฎที่ได้มานี้ยังต้องใช้การนับความถี่เข้าช่วยในกรณีที่อักษรเดียวกันในบริบทเดียวกันสามารถถอดอักษรได้หลายแบบ จึงใช้วิธีเก็บแบบการถอดอักษรที่พบมากที่สุดจะใส่ไว้ในตารางคาดเดาอักษร คือ เมื่อพบอักษรเป้าหมายตรงกับในตารางคาดเดานี้ ก็จะได้ผลการถอดอักษรในตารางนี้ซึ่งเป็นแบบที่พบมากที่สุดในคลังข้อมูลออกมา ตารางกฎของคอมพิวเตอร์ที่ได้ออกมาจะมีลักษณะที่เรียงลำดับจากตารางคาดเดาที่ไม่มีบริบทก่อน (g) จากนั้นจะเป็นตารางคาดเดาที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง(1g) แล้วจะเป็นตารางคาดเดาที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง (2g) ต่อไปจะเป็นตารางกฎที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง (1) และจะเป็นตารางกฎที่มีบริบทซ้าย 2 ตำแหน่งและขวา 3 ตำแหน่ง (2) สุดท้ายจะเป็นตารางกฎที่มีบริบทซ้าย 3 ตำแหน่งและขวา 3ตำแหน่ง (3) ดังแสดงให้เห็นคร่าวๆในรูป 4-1

ประเภท	อักษรอังกฤษ	
อักษรไทย		
g	rh	ร
g	ch	ค
g	ck	ก
g	gn	น
1g	b|r|e	ร
1g	m|m|^	ม์
1g	k|ai|i	ไ.
1g	u|ch|c	ช
2g	le|m|ma	ม
2g	^b|o|w^	โ.
2g	^s|p|od	ป
2g	co|n|ne	น
1	e|o|r	ออ
1	e|o|s	โอ
1	r|x|x	ช
1	m|t|a	ต
2	ib|r|ot	ร
2	cl|e|gh	เ.
2	ca|m|^^	ม
2	mo|l|ys	ล
3	omi|ng|g^^	ง
3	eru|s|sit	ส
3	mag|o|niu	โ.

รูป 4-1 แสดงตารางกฎสำหรับคอมพิวเตอร์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จะเห็นได้ว่าตารางกฎสำหรับคอมพิวเตอร์ด้านบนนั้นส่วนใหญ่แล้วจะมีบริบทซ้ายขวากำกับไว้อย่างละเอียดว่าอักษรภาษาอังกฤษเป้าหมายที่อยู่ระหว่างเครื่องหมาย "|" นั้นจะถอดอักษรเป็นอักษรภาษาไทยตัวใด หากไม่มีบริบทซ้ายขวาเลยเช่นตารางประเภท 'g' คือ ตารางคาดเดาที่ไม่มีบริบท ก็จะได้กฎการถอดอักษรโดยไม่มีบริบทนี้มาจากแบบการถอดอักษรที่มีความถี่มากที่สุดที่พบในคลังข้อมูล ตารางกฎสำหรับคอมพิวเตอร์ที่ประมวลผลได้ออกมานั้นจะมีจำนวนกฎที่พบในแต่ละตารางกฎดังนี้

1) ตารางกฎที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง มีจำนวนกฎ 2,743 กฎ

2) ตารางกฎที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง มีจำนวนกฎ 12,620 กฎ

3) ตารางกฎที่มีบริบทซ้าย 3 ตำแหน่งและขวา 3 ตำแหน่ง มีจำนวนกฎ 3,879 กฎ

4) ตารางคาดเดาที่ไม่มีบริบทมีจำนวนกฎ 82 กฎ

5) ตารางคาดเดาที่มีบริบทซ้าย 1 ตำแหน่งและขวา 1 ตำแหน่ง มีจำนวนกฎ 3,362 กฎ

6) ตารางคาดเดาที่มีบริบทซ้าย 2 ตำแหน่งและขวา 2 ตำแหน่ง มีจำนวนกฎ 9,019 กฎ

หากนำกฎที่ได้จากคอมพิวเตอร์มาเปรียบเทียบกับกฎการถอดอักษรของราชบัณฑิตยสถานก็จะเห็นได้ว่ากฎของราชบัณฑิตยสถานนั้น ไม่ได้ให้บริบทการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยละเอียดกับอักษรทุกตัว ที่เห็นได้ชัดจะเป็นส่วนของสระภาษาอังกฤษที่แทบจะไม่มีบริบทซ้ายขวากำหนดไว้เลย ดังจะเห็นได้จากตัวอย่างตารางเทียบเสียงสระภาษาอังกฤษดังตาราง 4-1

สระ	**ใช้**	**ตัวอย่าง**		
A				
a	แอ	b**a**dminton	=	แบดมินตัน
	อะ	**a**luminium	=	อะลูมิเนียม
	อา	Chic**a**go	=	ชิคาโก

	เอ	**A**sia	=	เอเชีย
	ออ	foot**a**ll	=	ฟุตบอล
ar	แอ	**ar**row	=	แอร์โรว์
E				
e	อี	Sw**e**den	=	สวีเดน
	เอ	L**e**banon	=	เลบานอน
	อิ	**e**lectronics	=	อิเล็กทรอนิกส์

ตาราง 4-1 แสดงตัวอย่างตารางเทียบเสียงสระภาษาอังกฤษของราชบัณฑิตยสถาน

จะเห็นได้ชัดว่าลักษณะการกำหนดกฎในการถอดอักษรภาษาอังกฤษเป็นภาษาไทยของโปรแกรมคอมพิวเตอร์กับกฎของราชบัณฑิตยสถานแตกต่างกัน ตารางกฎของคอมพิวเตอร์นั้นจะมีการระบุบริบทซ้ายขวาที่ละเอียดกว่ากฎของราชบัณฑิตยสถานเนื่องจากกฎของราชบัณฑิตยสถานนั้นเขียนขึ้นมาเพื่อให้คนนำกฎที่ให้ไปประยุกต์ใช้เอง แต่ตารางกฎที่กำหนดให้คอมพิวเตอร์นั้นมีปริมาณมากกว่าเพื่อที่จะให้คอมพิวเตอร์ประมวลผลข้อมูลที่รับเข้าไปให้ออกมาเป็นผลลัพธ์ที่ถูกต้องแม่นยำที่สุด

ลำดับต่อไปผู้วิจัยจะนำกฎที่กำหนดขึ้นจากฐานข้อมูลมาเปรียบเทียบความสอดคล้องกับกฎของราชบัณฑิตยสถาน โดยจะเปรียบเทียบแต่ละตัวอักษรภาษาอังกฤษ ว่าจากกฎการถอดอักษรที่ราชบัณฑิตยสถานบัญญัติไว้นั้น จากข้อมูลจริงได้พบว่าอักษรภาษาอังกฤษเป้าหมายถอดอักษรเป็นอักษรภาษาไทยตามที่ราชบัณฑิตยสถานวางไว้ร้อยละเท่าใด ผู้วิจัยมีสมมติฐานว่าตารางกฎที่สร้างโดยคอมพิวเตอร์นั้นจะสอดคล้องกับกฎที่ราชบัณฑิตยสถานวางไว้เพราะตารางกฎของคอมพิวเตอร์นั้นก็สร้างมาจากคลังข้อมูลคำทับศัพท์ที่ราชบัณฑิตยสถานพิมพ์เผยแพร่ไว้นั่นเอง ตารางความสอดคล้องนี้จะแสดงโดยเริ่มจากตารางเทียบสระภาษาอังกฤษแล้วจึงเป็นตารางเทียบพยัญชนะภาษาอังกฤษตามลำดับดังต่อไปนี้

4.1.1 ตารางเทียบความสอดคล้องของกฎการถอดสระภาษาอังกฤษเป็นภาษาไทยคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน

เมื่อเปรียบเทียบความสอดคล้องของการกฎถอดสระภาษาอังกฤษเป็นภาษาไทยของคอมพิวเตอร์กับกฎที่ราชบัณฑิตยสถานได้วางไว้จะได้ผลการเปรียบเทียบความสอดคล้องโดยสรุปตามลำดับความสอดคล้องของการถอดอักษรได้ดังนี้ หากต้องการดูตารางเปรียบเทียบโดยละเอียดสามารถดูได้ในภาคผนวก ก

สระภาษาอังกฤษ	ร้อยละของ ความสอดคล้อง (%)	ร้อยละของ ความไม่สอดคล้อง (%)
ar	100.00	0.00
are	100.00	0.00
aea	100.00	0.00
aer	100.00	0.00
air	100.00	0.00
aw	100.00	0.00
ayr	100.00	0.00
ear	100.00	0.00
eau	100.00	0.00
ee	100.00	0.00
eer	100.00	0.00
ei	100.00	0.00
eir	100.00	0.00
eur	100.00	0.00
ew	100.00	0.00
ire	100.00	0.00
ia	100.00	0.00
ier	100.00	0.00

iew	100.00	0.00
iu	100.00	0.00
ore	100.00	0.00
oa	100.00	0.00
oar	100.00	0.00
oi	100.00	0.00
oo	100.00	0.00
oor	100.00	0.00
ou	100.00	0.00
our	100.00	0.00
oy	100.00	0.00
ur	100.00	0.00
ure	100.00	0.00
ua	100.00	0.00
uir	100.00	0.00
i	99.77	0.23
u	99.49	0.51
or	99.29	0.71
y	99.08	0.92
er	98.91	1.09
au	98.72	1.28
ea	98.70	1.30
ie	98.61	1.39
ai	98.48	1.52
ion	97.87	2.13
a	97.72	2.28
o	97.70	2.30
ow	97.26	2.74

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ey	96.47	3.53
ae	94.44	5.56
ir	93.75	6.25
ay	91.67	8.33
eu	91.18	8.82
ere	90.91	9.09
aa	88.88	11.12
oe	87.80	12.20
ui	86.67	13.33
eo	85.42	14.58
uy	75.00	25.00
ao	66.67	33.33
oer	66.67	33.33
ye	57.14	42.86
e	53.79	46.21
yr	50.00	50.00
ue	40.00	60.00
eou	33.33	66.67
ค่าเฉลี่ยรวม	92.68	7.32

ตาราง 4-2 เปรียบเทียบความสอดคล้องการถอดสระ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางเทียบกฎการถอดสระภาษาอังกฤษเป็นภาษาไทยด้านบนจะเห็นได้ว่าอักษรที่ได้จากคลังข้อมูลของราชบัณฑิตยสถานที่นำมาสร้างเป็นกฎให้กับคอมพิวเตอร์นั้น จะมีค่าเฉลี่ยของความสอดคล้องโดยรวมร้อยละ 92.68 และมีค่าเฉลี่ยความไม่สอดคล้องโดยรวมร้อยละ 7.32 เมื่อนำผลของการเปรียบเทียบความสอดคล้องของกฎมาเรียงลำดับจากมากไปน้อย จะเห็นได้ว่าจากจำนวนสระทั้งหมด 64 ตัว มีสระที่มีการถอดอักษรสอดคล้องกันร้อยละ 100 จำนวน 33 อักษร คิดเป็นร้อยละ 51.56 ได้แก่ ar, are, aea, aer, air, aw, ayr, ear, eau, ee, eer, ei, eir, eur, ew,

ire, ia, ier, iew, iu, ore, oa, oar, oi, oo, oor, ou, our, oy, ur, ure, ua และ uir มีจำนวนสระที่มีการถอดอักษรสอดคล้องมากกว่าค่าเฉลี่ยจำนวนทั้งสิ้น 49 อักษร คิดเป็นร้อยละ 76.56 ส่วนสระที่มีความสอดคล้องของการถอดอักษรไม่ถึงค่าเฉลี่ยร้อยละ 92.68 มีจำนวน 15 อักษร คิดเป็นร้อยละ 23.44 ได้แก่ ay, eu, ere, aa, oe, ui, eo, uy, ao, oer, ye, e, yr, ue และ eou เป็นสระที่มีความสอดคล้องในการถอดอักษรน้อยที่สุดคือร้อยละ 33.33

ตัวอย่างของการถอดสระที่ไม่เป็นไปตามกฎการทับศัพท์ของราชบัณฑิตยสถานสามารถดูได้จากตาราง 4-3

สระภาษาอังกฤษ	**กฎการถอดอักษรของราชบัณฑิต**	**การถอดอักษรที่ไม่ตรงกับกฎของราชบัณฑิต**	**ตัวอย่าง**	**ถอดอักษรผิดจำนวน**	**คิดเป็นร้อยละ**
eou	โอ	เอีย	silic**eou**s=ซีลิเชียส	2	66.67
yr	เออ	อิ	p**yr**rhotite=พิร์โรไทต์	5	50.00
e	อี,เอ,อิ,เอะ	เออ	couple=คัปเปิล	80	3.19

ตาราง 4-3 ตัวอย่างการถอดสระที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน

จากตาราง 4-3 แสดงให้เห็นว่า eou ซึ่งควรจะถอดอักษรเป็น โอ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น เอีย 2 ครั้งจากจำนวนทั้งหมด 3 ครั้ง ส่วน yr ซึ่งควรจะถอดอักษรเป็น เออ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น เอีย 5 ครั้งจากจำนวนทั้งหมด 10 ครั้ง และ e ซึ่งควรจะถอดอักษรเป็น อี,เอ,อิ,เอะ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น เออ 80 ครั้งจากจำนวนทั้งหมด 1,349 ครั้ง

เมื่อเทียบความสอดคล้องของกฎการถอดสระแล้ว ในตารางถัดไปจะเทียบความสอดคล้องของกฎการถอดพยัญชนะ เพื่อดูว่ากฎที่ได้มาจากคอมพิวเตอร์นั้นสอดคล้องกันกับกฎที่ราชบัณฑิตยสถานให้ไว้เพียงไร

4.1.2 ตารางเทียบความสอดคล้องของกฎการถอดพยัญชนะภาษาอังกฤษเป็นภาษาไทยของคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน

เมื่อเทียบความสอดคล้องของกฎการถอดพยัญชนะภาษาอังกฤษเป็นภาษาไทยของคอมพิวเตอร์กับหลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานแล้ว แสดงได้เป็นตารางเปรียบเทียบความสอดคล้องโดยสรุปตามลำดับร้อยละของความสอดคล้อง แบ่งเป็นพยัญชนะต้นและพยัญชนะท้ายได้ดังนี้ หากต้องการดูรายละเอียดทั้งหมด สามารถดูได้ที่ภาคผนวก ข

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น	
	สอดคล้อง (%)	ไม่สอดคล้อง (%)
d	100	0
f	100	0
gn	100	0
h	100	0
k	100	0
kh	100	0
n	100	0
ph	100	0
q	100	0
rh	100	0
sch	100	0
v	100	0

wh	100	0
x	100	0
y	100	0
z	100	0
p	100	0
b	99.85	0.15
g+(a,o,u,l,r)	99.75	0.25
s	99.66	0.34
r	99.47	0.53
t	99.43	0.57
l	99.06	0.94
m	99.01	0.99
sh	98.96	1.04
ch	98.85	1.15
th	98.41	1.59
w	98.08	1.92
c+(a,o,u,l,r)	96.74	3.26
j	96.7	3.3
g+(e,i,y)	95.59	4.41
c+(e,l,y)	94.67	5.33
gh	85.71	14.29
sc	50	50
ค่าเฉลี่ยรวม	97.35	2.65

ตาราง 4-4 เปรียบเทียบความสอดคล้องการถอดพยัญชนะต้น

จากตาราง 4-4 ด้านบนจะเห็นได้ว่าอักษรที่ได้จากคลังข้อมูลของราชบัณฑิตยสถานที่นำมาสร้างเป็นกฎให้กับคอมพิวเตอร์นั้น จะมีค่าเฉลี่ยความสอดคล้องของพยัญชนะต้นโดยรวมร้อยละ 97.35 และมีค่าเฉลี่ยความไม่สอดคล้องโดยรวมร้อยละ 2.65 เมื่อนำผลของการ

เปรียบเทียบความสอดคล้องของกฎมาเรียงลำดับจากมากไปน้อย จะเห็นได้ว่าจากจำนวนพยัญชนะต้นทั้งหมด 34 ตัว มีพยัญชนะต้นที่มีการถอดอักษรสอดคล้องกันร้อยละ 100 จำนวน 17 อักษร คิดเป็นร้อยละ 50 ได้แก่ d, f, gn, h, k, kh, n, p, ph, q, rh, sch, v, wh, x, y และ z มีจำนวนพยัญชนะต้นที่มีการถอดอักษรสอดคล้องมากกว่าค่าเฉลี่ยจำนวนทั้งสิ้น 28 อักษร คิดเป็นร้อยละ 82.35 ส่วนพยัญชนะต้นที่มีความสอดคล้องของการถอดอักษรไม่ถึงค่าเฉลี่ยร้อยละ 97.35 มีจำนวน 6 อักษร คิดเป็นร้อยละ 17.65 ได้แก่ c+(a,o,u,l,r),j, g+(e,i,y), c+(e,i,y), gh และ sc เป็นพยัญชนะต้นที่มีความสอดคล้องในการถอดอักษรน้อยที่สุดคือร้อยละ 50

ตัวอย่างของการถอดพยัญชนะต้นที่ไม่เป็นไปตามกฎการทับศัพท์ของราชบัณฑิตยสถาน สามารถดูได้จากตาราง 4-5

พยัญชนะต้นภาษาอังกฤษ	**กฎการถอดอักษรของราชบัณฑิต**	**การถอดอักษรที่ไม่ตรงกับกฎของราชบัณฑิต**	**ตัวอย่าง**	**ถอดอักษรผิดจำนวน**	**คิดเป็นร้อยละ**
sc	ซ	ช	con**sc**ience=คอนเชียนซ์	1	50.00
gh	ก	ค	**gh**azal=คาซัล	1	14.29
ce	ซ	ค	**ce**ltic=เคลติก	3	1.23

ตาราง 4-5 ตัวอย่างการถอดพยัญชนะต้นที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน

จากตาราง 4-5 แสดงให้เห็นว่า sc ซึ่งควรจะถอดอักษรเป็น ซ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ช 1 ครั้งจากจำนวนทั้งหมด 2 ครั้ง ส่วน gh ซึ่งควรจะถอดอักษรเป็น ก ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ค 1 ครั้งจากจำนวนทั้งหมด 7 ครั้ง และ ce ซึ่งควรจะถอดอักษรเป็น ซ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ค 3 ครั้งจากจำนวนทั้งหมด 244 ครั้ง

เมื่อได้เห็นการเปรียบเทียบความสอดคล้องการถอดอักษรของพยัญชนะต้นไปแล้ว ลำดับต่อไปจะทำการเปรียบเทียบความสอดคล้องของการถอดพยัญชนะท้าย โดยจะสรุปให้เห็นตามลำดับร้อยละของความสอดคล้อง ดังดูได้จากตาราง 4-6

พยัญชนะภาษาอังกฤษ	พยัญชนะท้าย	
	สอดคล้อง (%)	ไม่สอดคล้อง (%)
b	100.00	0.00
ch	100.00	0.00
g	100.00	0.00
gh	100.00	0.00
gn	100.00	0.00
h	100.00	0.00
k	100.00	0.00
kh	100.00	0.00
n	100.00	0.00
ph	100.00	0.00
q	100.00	0.00
rh	100.00	0.00
sc	100.00	0.00
sh	100.00	0.00
sk	100.00	0.00
thm	100.00	0.00
v	100.00	0.00
y	100.00	0.00
c	99.42	0.58
m	99.21	0.79

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

d	98.92	1.08
r	98.81	1.19
th	97.78	2.22
x	97.37	2.63
p	96.87	3.13
t	96.42	3.58
l	95.82	4.18
f	95.24	4.76
s	92.94	7.06
z	91.67	8.33
c+(e)	91.49	8.51
g+(e)	89.66	10.34
sm	85.71	14.29
w	82.35	17.65
ค่าเฉลี่ยรวม	97.34	2.66

ตาราง 4-6 เปรียบเทียบความสอดคล้องการถอดพยัญชนะท้าย

จากตาราง 4-6 ด้านบนจะเห็นได้ว่าอักษรที่ได้จากคลังข้อมูลของราชบัณฑิตยสถานที่นำมาสร้างเป็นกฎให้กับคอมพิวเตอร์นั้น จะมีค่าเฉลี่ยความสอดคล้องของพยัญชนะท้ายโดยรวมร้อยละ 97.34 และมีค่าเฉลี่ยความไม่สอดคล้องโดยรวมร้อยละ 2.66 เมื่อนำผลของการเปรียบเทียบความสอดคล้องของกฎมาเรียงลำดับจากมากไปน้อย จะเห็นได้ว่าจากจำนวนพยัญชนะท้ายทั้งหมด 34 ตัว มีพยัญชนะท้ายที่มีการถอดอักษรสอดคล้องกันร้อยละ 100 จำนวน 18 อักษร คิดเป็นร้อยละ 52.94 ได้แก่ b, ch, g, gh, gn, h, k, kh, n, ph, q, rh, sc, sh, sk, thm, v และ y มีจำนวนพยัญชนะท้ายที่มีการถอดอักษรสอดคล้องมากกว่าค่าเฉลี่ยจำนวนทั้งสิ้น 24 อักษร คิดเป็นร้อยละ 70.59 ส่วนพยัญชนะท้ายที่มีความสอดคล้องของการถอดอักษรไม่ถึงค่าเฉลี่ยร้อยละ 97.34 มีจำนวน 10 อักษร คิดเป็นร้อยละ 29.41 ได้แก่ p, t, l, f, s, z, c+(e), g+(e), sm และ w เป็นพยัญชนะท้ายที่มีความสอดคล้องในการถอดอักษรน้อยที่สุดคือร้อยละ 82.35

ตัวอย่างของการถอดพยัญชนะท้ายที่ไม่เป็นไปตามกฎการทับศัพท์ของราชบัณฑิตยสถาน สามารถดูได้จากตาราง 4-7

พยัญชนะท้ายภาษาอังกฤษ	กฎการถอดอักษรของราชบัณฑิต	การถอดอักษรที่ไม่ตรงกับกฎของราชบัณฑิต	ตัวอย่าง	ถอดอักษรผิดจำนวน	คิดเป็นร้อยละ
sm	ซึม	สม	populism=ปอปปูลิสม์	1	4.76
ge	จ	ช	georges=ยอร์ช	3	10.34
ce	ซ	ส	venice=เวนิส	3	6.38

ตาราง 4-7 ตัวอย่างการถอดพยัญชนะท้ายที่ไม่เป็นไปตามกฎของราชบัณฑิตยสถาน

จากตาราง 4-7 แสดงให้เห็นว่า sm ซึ่งควรจะถอดอักษรเป็น ซึม ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น สม 1 ครั้งจากจำนวนทั้งหมด 21 ครั้ง ส่วน ge ซึ่งควรจะถอดอักษรเป็น จ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ช 3 ครั้งจากจำนวนทั้งหมด 204 ครั้ง และ ce ซึ่งควรจะถอดอักษรเป็น ซ ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ส 3 ครั้งจากจำนวนทั้งหมด 47 ครั้ง

เมื่อได้เห็นการเปรียบเทียบความสอดคล้องของสระ พยัญชนะต้น และพยัญชนะท้ายไปดังที่กล่าวไปข้างต้นแล้วนั้น จะเห็นได้ว่ามีอักษรจำนวนหนึ่งที่ถอดอักษรไม่สอดคล้องกับฎการถอดอักษรที่ราชบัณฑิตยสถานวางไว้ อักษรกลุ่มนี้สะท้อนให้เห็นว่าในคลังข้อมูลของราชบัณฑิตยสถานที่เก็บรวบรวมมานั้น มีคำทับศัพท์จำนวนหนึ่งที่ถอดอักษรไม่สอดคล้องกับกฎของราชบัณฑิตยสถาน เมื่อได้นำคลังข้อมูลที่จับคู่ข้อมูลไว้แล้ว มาเปรียบเทียบกับกฎของราชบัณฑิตยสถานพบว่ามีคำจำนวน 1,360 คำ จากจำนวนคำทั้งหมด 8,190 คำ ที่ถอดอักษรไม่

สอดคล้องกับกฎการถอดอักษรของราชบัณฑิตยสถาน คิดเป็นร้อยละ 16.61 สามารถแสดงให้เห็นเป็นตัวอย่างได้ในตาราง 4-8

คำภาษาอังกฤษ	**คำที่ถอดอักษรเป็นภาษาไทย**	**อักษรภาษาอังกฤษที่ถอดอักษรผิด**	**อักษรภาษาไทยที่ถอดอักษรผิด**	**การถอดอักษรตามกฎของราชบัณฑิต**
n/i/g/e/r/	น/ไ./จ/เ.อ/ร์/	ge	จ	ก
p/o/r/a/r/i/s/e/	พ/โ./ล/า/ร/ไ./ส์/#/	r	ล	ร
t/o/r/s/io/n/	ท/อ/ร์/ชั/น/	s	ช	ซ,ส
a/d/r/ia/n/o/p/l/e/	เอ/ด/ร/เ..ีย/น/โ./ป/ล/เ../	e	เ.อ	ี,,เ,,เ.ะ,
a/m/p/l/i/t/u/d/e/	แอ/ม/พ/ลิ//จู//ด/#/	t	จ	ท,ต,
c/l/a/p/h/a/m/	ค/ล/แ../ป/ฮ/แ./ม/	ph	ป	ฟ,
g/eo/r/g/e/	จ/อ/ร์/จ/#/	ge	จ	ก
h/e/l/i/c/	ฮี//ลิ//กซ์/	c	กซ	ช,ก
j/a/ck/	จ/แ็./ก/	a	แ็.	แ,ะ,า,เ,อ
l/a/s/z/l/o/	ล/า/ส/ส/#/โ./	z	ส	ซ
m/a/c/au/l/a/y/	ม/แ./กค/อ/ลี//ย์/	ay	ี	เ

ตาราง 4-8 ตัวอย่างคำที่ถอดอักษรไม่สอดคล้องกับกฎของราชบัณฑิตยสถาน

จากตาราง 4-8 แสดงให้เห็นว่าคำว่า niger นั้น ge ซึ่งควรจะถอดอักษรเป็น ก ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น จ ส่วนคำว่า porarise นั้น r ซึ่งควรจะถอดอักษรเป็น ร ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ล และคำว่า torsion นั้น s ซึ่งควรจะถอดอักษรเป็น ซ หรือ ส ตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้น พบว่ามีการถอดอักษรเป็น ช เป็นต้น

จากตารางการเปรียบเทียบความสอดคล้องการถอดสระและพยัญชนะในข้อ 4.1.1 และ 4.1.2 นั้นจะเห็นได้ว่ามีอักษรหลายตัวที่ถอดอักษรไม่สอดคล้องกับที่ราชบัณฑิตยสถานได้กำหนดหลักเกณฑ์เอาไว้ สาเหตุของความไม่สอดคล้องนี้สามารถสรุปได้เป็นสาเหตุหลักๆดังต่อไปนี้

4.1.3 สาเหตุของความไม่สอดคล้องของกฎที่ได้จากคอมพิวเตอร์และกฎของราชบัณฑิตยสถาน

ดังที่ได้กล่าวไปแล้วข้างต้นว่ามีการถอดอักษรที่ไม่สอดคล้องกับกฎที่ราชบัณฑิตยสถานวางไว้อยู่จำนวนหนึ่งนั้น หากต้องการดูข้อมูลที่ไม่สอดคล้องทั้งหมดสามารถดูได้จากภาคผนวก ก และ ข สำหรับตัวอย่างของความไม่สอดคล้องในการถอดอักษรบางส่วน โดยที่สาเหตุที่ทำให้เกิดความไม่สอดคล้องกันของการถอดอักษรที่ได้จากโปรแกรมคอมพิวเตอร์และการถอดอักษรของราชบัณฑิตยสถานนั้นอาจจะมาจากการที่กฎการถอดอักษรของราชบัณฑิตยสถานอาจไม่ครอบคลุมพอ นั่นคือ จากคลังข้อมูลที่เก็บมาจากคำที่ถอดอักษรโดยราชบัณฑิตยสถานนั้นพบว่ามีการถอดอักษรแบบอื่นๆ ที่นอกเหนือไปจากกฎการถอดอักษรที่ราชบัณฑิตยสถานให้ไว้จำนวนหนึ่ง ในการถอดอักษรแบบที่นอกเหนือจากกฎของราชบัณฑิตยสถานที่พบนั้นมีหลายแบบที่พบเป็นจำนวนมากและไม่ได้มีสาเหตุมาจากความหลากหลายในการออกเสียงด้วย ตัวอย่างเช่น การพบการถอดอักษร e เป็น เออ แทนที่จะเป็น อี, เอ และ อิ ตามที่ราชบัณฑิตยสถานให้กฎไว้ซึ่งพบจำนวน 80 ครั้ง จากจำนวนทั้งหมด 1349 ครั้ง เช่น คำว่า couple ที่ออกเสียงว่า /kʌpl/นั้นราชบัณฑิตยสถานได้ถอดอักษรไว้ในหลักเกณฑ์การทับศัพท์ภาษาอังกฤษว่า คัปเปิล แต่ไม่ได้มีการเขียนกฎที่ชัดเจนให้กับสระ e ที่ตามหลัง l ให้ถอดอักษรเป็น เออ ในคำว่า Georgia นั้นภาษาอังกฤษออกเสียงว่า /dʒɔrdʒə/ สระ eo ในคำนี้เมื่อเทียบกับภาษาไทยแล้วจะใกล้เคียงเสียงสระ ออ ซึ่งราชบัณฑิตยสถานยังไม่ได้เขียนกฎให้ครอบคลุมการถอดอักษรในคำนี้โดยกำหนดให้ eo ถอดอักษรได้เป็น อี, เอ, เอีย และเอียวเท่านั้น จากในข้อมูลพบการถอดอักษร eo เป็น ออ 7 ครั้งจากจำนวนทั้งหมด 41 ครั้ง สำหรับ s ท้ายคำนั้นพบว่า ไม่ถอดอักษรถึง 29 ครั้งจาก 425 ครั้ง ซึ่งราชบัณฑิตยสถานไม่ได้กำหนดเกณฑ์การไม่ถอดอักษร s ไว้ ตัวอย่างที่พบเช่น nabbes ตัวอย่างเหล่านี้เป็นข้อสังเกตว่ากฎการถอดอักษรที่ราชบัณฑิตยสถานวางไว้นั้นอาจจะยังไม่ครบถ้วนหรือครอบคลุมการถอดอักษรจริงเพียงพอ นอกจากนี้สาเหตุของความไม่สอดคล้องของของกฎอาจมีสาเหตุมาจากตัวภาษาอังกฤษเองที่เป็นภาษาที่มีการยืมคำจากภาษาอื่นเป็นจำนวนมากโดยเฉพาะภาษาฝรั่งเศส และภาษาลาติน จึงทำให้การออกเสียงคำภาษาอังกฤษหลายคำไม่

เป็นไปตามหลักการออกเสียงภาษาอังกฤษ อีกทั้งยังมีสาเหตุมากจากด้านประวัติศาสตร์ของภาษาด้วย นั่นคือ มีการออกเสียงคำในสมัยนั้นแล้วใช้กันจนติดในระดับคำแล้ว เช่น มีการออกเสียงตามครูในโรงเรียนฝรั่งที่ใช้สำเนียงภาษาตระกูลอินโดยูโรเปียน เช่น ฝรั่งเศส หรือ ปอร์ตุเกส จึงทำให้การออกเสียงคำภาษาอังกฤษผิดไปได้ ตัวอย่างของการถอดอักษรที่ไม่สอดคล้องสามารถดูตัวอย่างได้ในตาราง 4-9

อักษรภาษาอังกฤษ	การถอดอักษรของราชบัณฑิต	การถอดอักษรที่ไม่สอดคล้อง	ตัวอย่าง	ถอดอักษรไม่สอดคล้องจำนวน	คิดเป็นร้อยละ
e	อี, เอ, อิ, เอะ	เออ	cou**p**l**e**=คัปเปิล	80	3.19
eo	อี, เอ, เอีย, เอียว	ออ	g**eo**rgia=จอร์เจีย	7	14.58
s	ซ,ส,ช	#	nabbe**s**=แนบบ์	29	6.82

ตาราง 4-9 ตัวอย่างการถอดอักษรที่ไม่สอดคล้องเนื่องจากการพบการถอดอักษรที่นอกเหนือไปจากกฎการถอดอักษรของราชบัณฑิตยสถาน

หลังจากได้เปรียบเทียบความสอดคล้องของกฎการถอดอักษรของคอมพิวเตอร์และกฎการถอดอักษรของราชบัณฑิตยสถาน อีกทั้งได้เห็นสาเหตุของการถอดอักษรที่ไม่สอดคล้องดังที่ได้กล่าวไปแล้วข้างต้น ผู้วิจัยจะขอนำเสนอผลการทดลองถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานของโปรแกรมที่ได้พัฒนาขึ้นมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.2 ผลการทดลองถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน

ส่วนนี้จะกล่าวถึงผลการทำงานของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมาซึ่งได้นำไปทดลองถอดอักษรและทำการประเมินความถูกต้องดังที่ได้กล่าวมาแล้วในบทที่ 3 โดยในอันดับแรกจะนำเสนอผลการประเมินผลความถูกต้องแม่นยำในระดับคำของโปรแกรมการถอดอักษรที่พัฒนาขึ้นมากับคำที่ราชบัณฑิตยสถานถอดอักษรไว้ จากจำนวนคำทั้งหมด 8,190 คำ เมื่อนำมาทดสอบประสิทธิภาพการทำงานของโปรแกรมจะได้ผลออกมาดังดูได้จากตาราง 4-10

การทดสอบครั้งที่	จำนวนคำทั้งหมด (คำ)	จำนวนคำที่ถอดอักษรถูกต้อง (คำ)	จำนวนคำที่ถอดอักษรผิดพลาด (คำ)	ร้อยละของความถูกต้องแม่นยำ (%)
1	1,637	622	1,015	38.00
2	1,639	639	1,000	38.99
3	1,638	644	994	39.32
4	1,638	682	956	41.64
5	1,638	650	988	39.68
เฉลี่ย	1,638	647.4	991	39.53

ตาราง 4-10 ประสิทธิภาพในการถอดอักษรของโปรแกรมการถอดอักษรในระดับคำ

จากตารางข้างต้นนั้นจะเห็นได้ว่าในแต่ละครั้งนั้น โปรแกรมถอดอักษรได้ผลความถูกต้องใกล้เคียงกัน โดยที่ในการทดสอบครั้งแรกซึ่งมีข้อมูลชุดที่ 1, 2, 3 และ 4 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 5 จำนวน 1,638 คำ เป็นคลังข้อมูลทดสอบนั้น ถอดอักษรได้ถูกต้อง 622 คำ และผิดพลาด 1,015 คำ คิดเป็นความถูกต้องร้อยละ 38.00 ในการทดสอบครั้งที่ 2 มีข้อมูลชุดที่ 2,3,4 และ 5 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 1 จำนวน 1,639 คำ เป็นคลังข้อมูลทดสอบนั้น ถอดอักษรได้ถูกต้อง 639 คำ และผิดพลาด 1,000 คำ คิดเป็นความถูกต้องร้อยละ 38.99 ในการทดสอบครั้งที่ 3 มีข้อมูลชุดที่ 3,4,5 และ 1 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 2 จำนวน

1,638 คำ เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 644 คำ และผิดพลาด 994 คำ คิดเป็นความถูกต้องร้อยละ 39.32 ในการทดสอบครั้งที่ 4 มีข้อมูลชุดที่ 4,5,1 และ 2 เป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 3 จำนวน 1,638 คำ เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 682 คำ และผิดพลาด 956 คำ คิดเป็นความถูกต้องร้อยละ 41.64 สำหรับในการทดสอบครั้งสุดท้ายคือครั้งที่ 5 มีข้อมูลชุดที่ 5,1,2 และ 3 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 4 จำนวน 1,638 คำ เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 650 คำ และผิดพลาด 988 คำ คิดเป็นความถูกต้องร้อยละ 39.68 เมื่อนำร้อยละของความถูกต้องในแต่ละครั้งมาเฉลี่ยกันแล้ว จะได้ค่าเฉลี่ยความถูกต้องแม่นยำร้อยละ 39.53

จากการทดลองการถอดอักษรนี้จะเห็นได้ว่าคำที่ถอดอักษรถูกต้องนั้นมีน้อยกว่าคำที่ถอดอักษรผิดพลาดจึงทำค่าเฉลี่ยความถูกต้องจากการทดสอบการถอดอักษรทั้ง 5 ครั้งนั้นไม่สูงนัก ผู้วิจัยจึงนำโปรแกรมมาทดสอบประสิทธิภาพความถูกต้องแม่นยำในระดับตัวอักษรอีกครั้ง เพื่อจะดูผลที่ได้ว่ามากหรือน้อยกว่าความแม่นยำในระดับคำเพียงใด จึงได้ผลออกมาดังตาราง 4-11

การทดสอบครั้งที่	**จำนวนอักษรทั้งหมด (อักษร)**	**จำนวนอักษรที่ถอดอักษรถูกต้อง (อักษร)**	**จำนวนอักษรที่ถอดอักษรผิดพลาด (อักษร)**	**ร้อยละของความถูกต้องแม่นยำ (%)**
1	11,398	9,029	2,369	79
2	11,285	8,995	2,290	80
3	11,263	8,976	2,287	80
4	11,113	8,883	2,230	80
5	11,301	9,049	2,252	80
เฉลี่ย	11,272	8,986.4	2,285.6	79.8

ตาราง 4-11 ประสิทธิภาพในการถอดอักษรของโปรแกรมการถอดอักษรในระดับอักษร

จากตาราง 4-11 จะเห็นได้ว่าในแต่ละครั้งนั้น โปรแกรมถอดอักษรได้ผลความถูกต้องใกล้เคียงกัน โดยที่ในการทดสอบครั้งแรกซึ่งมีข้อมูลชุดที่ 1, 2, 3 และ 4 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 5 จำนวน 11,398 อักษร เป็นคลังข้อมูลทดสอบนั้น ถอดอักษรได้ถูกต้อง 9,029

อักษรและผิดพลาด 2,369 อักษร คิดเป็นความถูกต้องร้อยละ 79 ในการทดสอบครั้งที่ 2 มีข้อมูลชุดที่ 2,3,4 และ 5 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 1 จำนวน 11,285 อักษร เป็นคลังข้อมูลทดสอบนั้น ถอดอักษรได้ถูกต้อง 8,995 อักษรและผิดพลาด 2,290 อักษร คิดเป็นความถูกต้องร้อยละ 80 ในการทดสอบครั้งที่ 3 มีข้อมูลชุดที่ 3,4,5 และ 1 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 2 จำนวน 11,263 อักษร เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 8,976 อักษรและผิดพลาด 2,287 อักษรคิดเป็นความถูกต้องร้อยละ 80 ในการทดสอบครั้งที่ 4 มีข้อมูลชุดที่ 4,5,1 และ 2 เป็นคลังข้อมูลฝึกสอน และใช้ข้อมูลชุดที่ 3 จำนวน 11,113 อักษร เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 8,883 อักษรและผิดพลาด 2,230 อักษรคิดเป็นความถูกต้องร้อยละ 80 สำหรับในการทดสอบครั้งสุดท้ายคือครั้งที่ 5 มีข้อมูลชุดที่ 5,1,2 และ 3 เป็นคลังข้อมูลฝึกสอน และมีข้อมูลชุดที่ 4 จำนวน 11,301 อักษร เป็นคลังข้อมูลทดสอบ ถอดอักษรได้ถูกต้อง 9,049 อักษร และผิดพลาด 2,252 อักษร คิดเป็นความถูกต้องร้อยละ 80 เมื่อนำร้อยละของความถูกต้องในแต่ละครั้งมาเฉลี่ยกันแล้ว จะได้ค่าเฉลี่ยความถูกต้องแม่นยำร้อยละ 79.8

ดังนั้นจะเห็นได้ชัดว่าเมื่อทดสอบประสิทธิภาพความแม่นยำของโปรแกรมในระดับอักษรจะมีความแม่นยำมากกว่าระดับคำถึงร้อยละ 40.27 ทั้งนี้เป็นเพราะในระดับคำนั้นจะประกอบด้วยอักษรหลายตัว หากอักษรตัวใดตัวหนึ่งผิดก็จะทำให้คำทั้งคำถูกพิจารณาว่าถอดอักษรผิดทั้งคำ ความแม่นยำของการถอดอักษรในระดับอักษรจึงมีมากกว่า สำหรับตัวอย่างของผลของการถอดอักษรที่ถูกต้องจะดูได้จากตาราง 4-12 ส่วนตัวอย่างของผลการถอดอักษรที่ผิดพลาดนั้นสามารถดูได้จากตาราง 4-13 ดังนี้

คำภาษาอังกฤษ	**คำที่ถอดโดย โปรแกรมถอดอักษร**	**คำที่ถอดโดย ราชบัณฑิตยสถาน**
abney	แอบนีย์	แอบนีย์
acrylic	อะคริลิก	อะคริลิก
actin	แอกทิน	แอกทิน
acton	แอกตัน	แอกตัน
adolf	อดอล์ฟ	อดอล์ฟ
africa	แอฟริกา	แอฟริกา
agnew	แอกนิว	แอกนิว
albee	แอลบี	แอลบี

alkali	แอลคาไล	แอลคาไล
america	อเมริกา	อเมริกา

ตาราง 4-12 ตัวอย่างคำที่ถอดอักษรถูกต้อง

คำภาษาอังกฤษ	**คำที่ถอดโดย โปรแกรมถอดอักษร**	**คำที่ถอดโดย ราชบัณฑิตยสถาน**
abaco	อาบาโค	แอบาโก
abbas	อับเบส	อับบาส
abbott	อับโบตต์	แอบบอตต์
abdal	อับดอล	อับดัล
abelian	อาบีลิาน	อาบีเลียน
abraham	แอบราแฮม	เอบราแฮม
absorbance	แอบเสอร์แบนซ์	แอบซอร์แบนซ์
abyssinia	แอบิสซิไนา	อะบิสซิเนีย
acacia	อะกาซิา	อะเคเชีย
acanthite	อะแกนตฮิต์	อาแคนไทต์

ตาราง 4-13 ตัวอย่างคำที่ถอดอักษรผิดพลาด

จากตัวอย่างของคำที่ถอดอักษรผิดพลาดในตาราง 4-13 จะเห็นได้ว่าคำที่ถอดอักษรผิดส่วนใหญ่นั้น จะถอดอักษรในส่วนของสระผิด เช่น คำว่า abaco ถอดอักษร สระแอ เป็นสระอา และคำว่า abbas ถอดอักษร สระอา เป็น สระเอ นอกจากนี้ก็มีการถอดพยัญชนะผิดด้วย เช่น คำว่า absorbance ถอดพยัญชนะ ‘ซ’ เป็น ‘ส’ และคำว่า acacia ถอดพยัญชนะ ‘ค’ เป็น ‘ก’ จากคำที่ถอดอักษรผิดพลาดเหล่านี้ สามารถสรุปสาเหตุของการถอดอักษรที่ผิดพลาดของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานได้ดังต่อไปนี้

4.2.1 สาเหตุที่ทำให้การถอดอักษรโดยโปรแกรมการถอดอักษรผิดพลาด

ดังที่ได้กล่าวไปแล้วข้างต้นว่าโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นนี้ อาศัยคลังคำทับศัพท์ของราชบัณฑิตยสถานในการสร้างกฎการถอดอักษรให้กับคอมพิวเตอร์ได้เรียนรู้และประมวลผลการถอดอักษรออกมา ในเบื้องต้นนั้น ผู้วิจัยมีสมมติฐานว่าการเก็บรวบรวมคำทับศัพท์ที่ราชบัณฑิตยสถานพิมพ์เผยแพร่ไว้ จะสามารถช่วยสร้างกฎการถอดอักษรที่ถูกต้องได้ ทั้งนี้ผู้วิจัยมีความคาดหมายว่าคำที่ราชบัณฑิตถอดอักษรไว้นั้นจะถูกต้องตามหลักเกณฑ์การทับศัพท์ที่ราชบัณฑิตยสถานได้วางไว้ แต่เมื่อได้ทำการเก็บรวบรวมข้อมูลจนได้ตามกำหนดแล้ว และนำข้อมูลที่เก็บรวบรวมมาสร้างเป็นกฎในการถอดอักษรให้กับคอมพิวเตอร์ ก็พบว่าประสิทธิภาพในการถอดอักษรของโปรแกรมทำได้ไม่ดีนัก ทั้งนี้เมื่อวิเคราะห์ขั้นตอนการดำเนินการวิจัยต่างๆและข้อมูลที่มีการถอดอักษรโดยโปรแกรมผิดพลาดแล้วนั้น สามารถสรุปสาเหตุที่ทำให้โปรแกรมถอดอักษรมีความถูกต้องแม่นยำไม่มากได้ดังนี้

4.2.1.1 การถอดอักษรในคลังข้อมูลไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถานเสมอ

เมื่อวิเคราะห์คลังข้อมูลที่เก็บรวบรวมมาแล้วนั้น จะเห็นได้ว่ามีการถอดอักษรของคำหลายคำในคลังข้อมูลที่ไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน ทั้งนี้เป็นเพราะคลังข้อมูลที่เก็บรวบรวมมานั้นได้รับการถอดอักษรจากคนหลายคน จึงทำให้เกิดความหลากหลายในการถอดอักษร ซึ่งในบางครั้งการถอดอักษรนั้นไม่เป็นไปตามกฎเกณฑ์การทับศัพท์ที่ราชบัณฑิตยสถานได้วางเอาไว้ จึงส่งผลให้กฎการถอดอักษรของโปรแกรมมีประสิทธิภาพด้อยลงไป ตัวอย่างของการถอดอักษรในคลังข้อมูลที่ไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน เช่น

คำภาษาอังกฤษ	คำภาษาไทย
b/l/a/ck/	บ/ล/แ็../ก/
f/l/o/p/s/	ฟ/ล็/อ/ป/ส์/
l/i/n/d/g/r/e/n/	ล/ิ/น/-/ก/ร/เ็../น/

ตาราง 4-14 ตัวอย่างการถอดอักษรในคลังข้อมูลที่ไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน

จากตาราง 4-14 นั้น คำว่า ‘black’ ถอดอักษรเป็น ‘แบล็ก’ นั้นไม่เป็นไปตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถานเพราะราชบัณฑิตยสถานกำหนดว่า ‘a’ ให้ถอดอักษรเป็น ‘แอ,อะ,อา, เอ,ออ, ไม่ต้องใช้ไม้ไต่คู้เพราะไม้ไต่คู่นั้นจะใช้เพื่อให้เห็นความแตกต่างจากคำไทยและเพื่อช่วยให้ผู้อ่านแยกพยางค์ได้ถูกต้องเท่านั้น ส่วนคำว่า ‘flops’ ที่ถอดอักษรเป็น ‘ฟล็อปส์’ นั้น ราชบัณฑิตยสถานกำหนดว่าสระ ‘o’ ให้ถอดอักษรเป็น ‘ออ’ เลยไม่ต้องใช้ไม้ไต่คู้ ในส่วนของไม้ไต่คู่นั้นจะใช้เพื่อใช้เห็นความแตกต่างจากคำไทยและเพื่อช่วยให้ผู้อ่านแยกพยางค์ได้ถูกต้องเท่านั้น สำหรับคำว่า ‘Lindgren’ ที่ถอดอักษรเป็น ‘ลินเกร็น’ นั้นราชบัณฑิตยสถานกำหนดให้อักษรที่ไม่ออกเสียงให้ใส่เครื่องหมายทัณฑฆาต ฉะนั้นอักษร ‘d’ ที่ไม่ออกเสียง ต้องถอดอักษรเป็น ‘ด์’ และไม้ไต่คู้ก็ไม่จำเป็นต้องใช้ตามกฎที่ได้กล่าวไปแล้ว (หากต้องการดูหลักเกณฑ์การทับศัพท์ภาษาอังกฤษของราชบัณฑิตยสถานเพิ่มเติม สามารถดูได้ในบทที่ 2) สำหรับรายการคำที่ถอดอักษรไม่ตรงตามหลักเกณฑ์ของราชบัณฑิตยสถานนั้นมีจำนวนทั้งสิ้น 1,360 คำ จากจำนวนคำทั้งหมด 8,190 คำ ที่ถอดอักษรไม่สอดคล้องกับกฎการถอดอักษรของราชบัณฑิตยสถาน คิดเป็นร้อยละ 16.61 สามารถดูตัวอย่างเพิ่มเติมได้ในตาราง 4-8

4.2.1.2 **ผู้ถอดอักษรในคลังข้อมูลออกเสียงคำภาษาอังกฤษแตกต่างกัน**

ในส่วนของสาเหตุที่ผู้ถอดอักษรในคลังข้อมูลคำทับศัพท์ของราชบัณฑิตยสถานออกเสียงคำภาษาอังกฤษแตกต่างกันนั้น อาจสืบเนื่องมาจากหลายสาเหตุ ประการแรกเป็นเพราะภาษาอังกฤษนั้นเป็นภาษาที่มีการยืมคำภาษาต่างประเทศเป็นจำนวนมาก โดยที่คำภาษาอังกฤษจำนวนมากนั้นมาจากภาษาฝรั่งเศสในช่วงศตวรรษที่ 9 ภาษาฝรั่งเศสเป็นภาษาตระกูลอินโดยูโรเปียน แตกต่างกับภาษาอังกฤษที่อยู่ในตระกูลภาษาแองโกลแซกซอน ดังนั้นจึงมีอิทธิพลทำให้ภาษาอังกฤษหลายคำออกเสียงแตกต่างออกไป นอกจากนี้ยังมีสาเหตุมาจากทางด้านประวัติศาสตร์ เช่น การที่มีการออกเสียงคำภาษาอังกฤษจนติดมาตั้งแต่สมัยก่อน โดยอาจจะออกเสียงตามครูในโรงเรียนฝรั่งที่ใช้ภาษาตระกูลอินโดยูโรเปียน เช่น ฝรั่งเศส หรือ โปรตุเกส จึงทำให้มีการออกเสียงที่แตกต่างไปจากเจ้าของภาษาที่พูดอังกฤษเป็นภาษาแม่ได้ อีกทั้งยังอาจมีสาเหตุมาจากการที่ผู้ถอดอักษรคุ้นเคยกับสำเนียงอังกฤษแบบอังกฤษหรือสำเนียงอังกฤษแบบอเมริกันมากกว่า และเนื่องจากผู้ถอดอักษรไม่ได้ใช้ภาษาอังกฤษเป็นภาษาแม่ จึงอาจทำให้เกิดความสับสนในการอ่านออกเสียงคำภาษาอังกฤษได้ อีกทั้งเมื่อผู้ถอดอักษรมีจำนวนหลายคน และแต่ละคนอาจจะมีวิธีการออกเสียงคำแต่ละคำแตกต่างกันไป จึงสามารถส่งผลให้กฎการถอดอักษรของ

คอมพิวเตอร์ไม่แม่นยำได้ ตัวอย่างการที่ผู้ถอดอักษรในคลังข้อมูลคำทับศัพท์ของราชบัณฑิตยสถานออกเสียงคำภาษาอังกฤษแตกต่างกันนั้น จะเห็นได้ชัดเจนจากคำภาษาอังกฤษที่มีการถอดอักษรเป็นภาษาไทยหลายแบบ ทั้งที่อาจจะเป็นหนังสือเล่มเดียวกันหรือต่างเล่มกัน แม้ว่าการถอดอักษรคำเดียวกันเป็นหลายแบบนี้ ผู้วิจัยได้ทำการคัดเลือกแบบการถอดอักษรที่เป็นไปตามมาตรฐานของราชบัณฑิตยสถานมากที่สุดเพียงแบบเดียวไปแล้ว แต่ตัวอย่างในกลุ่มนี้แสดงให้เห็นได้ชัดถึงการออกเสียงที่แตกต่างกันของผู้ถอดอักษร จึงนำมาเสนอให้เห็นเป็นตัวอย่าง เช่น

คำภาษาอังกฤษ	**คำภาษาไทย**
l/eo/n/a/r/d/	ล/เ./น/า/ร์/ด/
l/eo/n/a/r/d/	ล/เ./นน/เ./ร์/ด/
l/eo/n/a/r/d/	ล/เ.ียว/น/า/ร์/ด/

ตาราง 4-15 ตัวอย่างข้อมูลที่แสดงว่าผู้ถอดอักษรในคลังข้อมูลคำทับศัพท์ของราชบัณฑิตยสถานออกเสียงคำภาษาอังกฤษแตกต่างกัน

จากตาราง 4-15 จะเห็นได้ว่าคำภาษาอังกฤษเพียงคำเดียวคือคำว่า ‘Leonard’ นั้นมีผู้ถอดอักษรได้ถึง 3 แบบ ซึ่งแต่ละแบบนั้น ก็เป็นไปตามเกณฑ์การถอดอักษรของราชบัณฑิตยสถาน จึงสะท้อนให้เห็นถึงการอ่านคำๆนี้ที่แตกต่างกัน ฉะนั้นคำเป็นจำนวนมากในคลังข้อมูลที่มีผู้ถอดอักษรที่แตกต่างกันก็อาจมีการออกเสียงที่หลากหลาย ทำให้กฎการถอดอักษรของโปรแกรมมีความคลาดเคลื่อนได้

4.2.1.3 **คลังข้อมูลไม่ได้มีการกำหนดขอบเขตของพยางค์**

ในตอนแรกนั้นผู้วิจัยไม่ได้ให้ความสำคัญในการกำหนดขอบเขตพยางค์ให้กับคลังข้อมูล เนื่องจากเห็นว่าการเขียนโปรแกรมสร้างกฎโดยดูจากบริบทซ้ายขวาก็น่าจะเพียงพอที่จะสร้างกฎการถอดอักษรที่แม่นยำได้ แต่เมื่อได้เห็นผลการถอดอักษรของโปรแกรมการถอดอักษรแล้วพบว่าการถอดอักษรของโปรแกรมที่ผิดพลาดหลายคำนั้นมีสาเหตุเนื่องมาจากการที่ไม่ได้กำหนดขอบเขตพยางค์ไว้ ทำให้คอมพิวเตอร์เกิดความสับสนว่าอักษรใดเป็นพยัญชนะต้น

หรือพยัญชนะท้ายโดยเฉพาะเมื่ออักษรนั้นอยู่ในคำหลายพยางค์และไม่ได้เป็นพยางค์แรกและพยางค์สุดท้ายของคำจึงทำให้เกิดความผิดพลาดในการถอดอักษรได้ดังตัวอย่าง ในตาราง 4-15

คำภาษาอังกฤษ	**คำที่ถอดโดย โปรแกรมถอดอักษร**	**คำที่ถอดโดย ราชบัณฑิตยสถาน**
coccolithophore	ค็อกโคลิตโฮพฮอร์	ค็อกโคลิโทฟอร์
congregation	คองกรีเกชไอออน	คองกรีเกชัน
eratosthenes	อิเรโตสตเฮนีส์	เอราทอสเทนีส

ตาราง 4-16 ตัวอย่างข้อมูลที่ถอดอักษรผิดเนื่องจากไม่ได้กำหนดขอบเขตพยางค์

จะเห็นได้ว่าคำว่า coccolithophore นั้นโปรแกรมจะถอดอักษร ‘t’ เป็น ‘ต’ ให้เป็นพยัญชนะท้ายของพยางค์ ‘ลิต’ และถอดอักษร ‘h’ เป็น ‘ฮ’ ซึ่งเป็นพยัญชนะต้นของพยางค์ ‘โฮ’ และถอดอักษร ‘p’ เป็น ‘พ’ และเป็นพยัญชนะท้ายของพยางค์ ‘โฮพ’ อีกทั้งถอดอักษร ‘h’ เป็น ‘ฮ’ และให้เป็นพยัญชนะต้นของพยางค์ ‘ฮอร์’ ซึ่งเป็นการถอดอักษรที่ผิด โดยที่ควรถอด ‘th’ เป็น ‘ท’ และ ‘ph’ เป็น ‘ฟ’ สำหรับคำว่า congregation นั้น โปรแกรมถอดอักษรโดยมอง ‘tion’ แยกจากกันเป็น ‘ชไอออน’ ซึ่งสืบเนื่องมาจากการที่โปรแกรมไม่ได้นำขอบเขตของพยางค์มาพิจารณาด้วย ตัวอย่างคำสุดท้ายคือคำว่า Eratosthenes นั้นเป็นกรณีเดียวกับคำแรกคือคอมพิวเตอร์มองอักษร ‘th’ แยกจากกัน โดยถอดอักษรเป็น ‘ต’ และ ‘ฮ’ ตามลำดับ ตัวอย่างในตารางข้างบนนี้สืบเนื่องมาจากการที่ไม่ได้กำหนดขอบเขตพยางค์ให้กับข้อมูลในคลังข้อมูลที่รวบรวมไว้

4.2.1.4 การเลือกอักษรในการถอดอักษรที่ผิดของโปรแกรม

จากผลการถอดอักษรโดยโปรแกรมการถอดอักษรจำนวนมากแสดงให้เห็นว่าโปรแกรมเลือกอักษรที่จะนำมาเป็นผลลัพธ์ของการถอดอักษรผิด ทั้งนี้อาจเป็นเพราะในบริบทเดียวกันตัวอักษรนั้นถอดได้หลายแบบ ทำให้คอมพิวเตอร์ไม่สามารถสรุปกฎลงในตารางกฎได้ ต้องใช้แบบการถอดอักษรที่มีความถี่มากที่สุดในการแก้ปัญหาการถอดอักษรที่มีบริบทเดียวกันแทน ซึ่งจะส่งผลให้คำจำนวนหนึ่งถอดอักษรผิดพลาดได้ การถอดอักษรได้หลายแบบนั้นจึงทำให้

ได้ผลการถอดอักษรที่ไม่ถูกต้องแม่นยำเพียงพอ ตัวอย่างการเลือกอักษรมาเป็นผลลัพธ์ของการถอดอักษรผิด เช่น

คำภาษาอังกฤษ	คำที่ถอดโดยโปรแกรมถอดอักษร	คำที่ถอดโดยราชบัณฑิตยสถาน
acute	แอคัต	อะคิวต์
albany	แอลบานี	ออลบานี
bengal	เบนกาล	เบงกอล

ตาราง 4-17 ตัวอย่างคำที่ถอดอักษรผิดเพราะโปรแกรมเลือกอักษรที่จะนำมาเป็นผลลัพธ์ของการถอดอักษรผิด

จากตาราง 4-17 จะเห็นได้ว่าคำว่า acute นั้นโปรแกรมเลือกแบบการถอดอักษรเป็น ‘แอคัต’ แทนที่จะเป็น ‘อะคิวต์’ เพราะโปรแกรมใช้ตารางคาดเดาที่คัดเลือกแบบการถอดอักษรที่พบมากที่สุดในคลังข้อมูล จึงถอดอักษร ‘a’ เป็น ‘แอ’ และ ‘u’ เป็น ‘อะ’ ส่วนคำว่า Albany นั้น โปรแกรมเลือกถอดอักษร ‘al’ เป็น ‘แอล’ แทนที่จะเป็น ‘ออล’ สำหรับคำว่า bangal นั้น โปรแกรมเลือกการถอดอักษรให้เป็น ‘เบนกาล’ แทนที่จะเป็น ‘เบงกอล’ การเลือกอักษรมาเป็นผลลัพธ์ในการถอดอักษรที่ผิดทำให้ประสิทธิภาพการถอดอักษรของโปรแกรมไม่ดีเท่าที่ควร ทางแก้ไขอันหนึ่งคือขยายขอบเขตบริบทให้มากกว่า 2 ตัวอักษร

4.2.1.5 การไม่มีรูปแทนเสียงเป็นตัวกลางในการถอดอักษร

จากผลการถอดอักษรโดยโปรแกรม จะเห็นได้ว่าผลการถอดอักษรที่ผิดพลาดส่วนหนึ่งมีผลมาจากการที่การถอดอักษรไม่มีรูปแทนเสียงภาษาอังกฤษเป็นตัวกลางในการถอดอักษร จึงทำให้เกิดความหลากหลายในการถอดอักษรขึ้น ส่งผลให้ประสิทธิภาพของโปรแกรมการถอดอักษรไม่ดีเท่าที่ควร ตัวอย่างคำที่ถอดอักษรผิดเพราะไม่มีรูปแทนเสียงเป็นตัวกลางในการถอดอักษรสามารถดูได้จากตาราง 4-18

คำ ภาษาอังกฤษ	คำที่ถอดโดย โปรแกรมถอดอักษร	คำที่ถอดโดย ราชบัณฑิตยสถาน
hamitic	แฮไมติก	แฮมิติก
maxwell	มากซ์เวลล์	แมกซ์เวลล์
pearl	แพร์ล	เพิร์ล

ตาราง 4-18 ตัวอย่างคำที่ถอดอักษรผิดเพราะการไม่มีรูปแทนเสียงเป็นตัวกลางในการถอดอักษร

จากตารางข้างต้นจะเห็นได้ว่าคำว่า ‘hamitic’ นั้น โปรแกรมเลือกถอดอักษร ‘i’ เป็น ‘ไอ’ แทนที่จะเป็น ‘อิ’ ในคำว่า ‘maxwell’ นั้น ถอดอักษร ‘a’ เป็น ‘อา’ แทนที่จะเป็น ‘แอ’ และในคำว่า ‘pearl’ ถอดอักษร ‘ea’ เป็น ‘แอ’ แทนที่จะเป็น ‘เออ’ หากมีรูปแทนเสียงเป็นตัวกลางในการถอดอักษรก็จะทำให้การถอดอักษรถูกต้องแม่นยำมากยิ่งขึ้น

เมื่อได้เห็นผลการทดลองถอดอักษรของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานไปดังที่กล่าวไปแล้วข้างต้นนั้น จะเห็นได้ว่าประสิทธิภาพในการถอดอักษรของโปรแกรมร้อยละ 39.53 นั้นยังไม่ดีนัก ทั้งนี้อาจมาจากสาเหตุ 5 ข้อดังที่กล่าวไปแล้ว หากต้องการพัฒนาโปรแกรมการถอดอักษรให้ดียิ่งขึ้น จำเป็นต้องเปลี่ยนขั้นตอนการดำเนินการวิจัยในหลายขั้นตอน รวมทั้งอาจต้องมีการเสริมขั้นตอนบางขั้นตอนขึ้นมาเพื่อให้ได้โปรแกรมการถอดอักษรที่มีประสิทธิภาพมากยิ่งขึ้น

ในบทต่อไปจะกล่าวสรุปผลการทำงานของโปรแกรมการถอดอักษรและให้ข้อเสนอแนะเพื่อเป็นแนวทางที่จะนำไปพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปและข้อเสนอแนะ

บทสุดท้ายนี้จะกล่าวสรุปกระบวนการและผลการทำงานของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน และนำมาตรวจสอบกับสมมติฐานที่ตั้งไว้ก่อนหน้าเพื่อดูว่า การศึกษาในวิทยานิพนธ์เล่มนี้ได้ผลลัพธ์เป็นไปตามที่คาดไว้หรือไม่ อย่างไร นอกจากนี้จะให้ข้อเสนอแนะเกี่ยวกับแนวทางที่จะพัฒนาปรับปรุงโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยต่อไป

5.1 สรุปกระบวนการในการพัฒนาโปรแกรมการถอดอักษรและผลการทำงานของโปรแกรมการถอดอักษร

วิทยานิพนธ์เล่มนี้ได้พัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถาน เพื่อให้ได้โปรแกรมการถอดอักษรที่มีกฎการถอดอักษรเช่นเดียวกับที่ราชบัณฑิตยสถานวางไว้ ทำให้การถอดอักษรตามมาตรฐานที่ราชบัณฑิตยสถานวางไว้เป็นไปได้ง่ายและสะดวกขึ้น ผู้วิจัยได้มีการประยุกต์ใช้แบบจำลอง(Model) แบบเปรียบเทียบตารางกฎ (Table Lookup) ตามแนวของ Van den Bosch and Daelemans (1993) ที่ใช้วิธีนำข้อมูลที่เก็บรวบรวมมาสร้างเป็นตารางกฎ เพื่อให้คอมพิวเตอร์นำไปประมวลผลในการถอดอักษร ผู้วิจัยมีสมมติฐานว่าคลังข้อมูลคำทับศัพท์ที่พิมพ์เผยแพร่โดยราชบัณฑิตยสถานนั้น จะสามารถสร้างกฎการถอดอักษรให้กับโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่มีความถูกต้องแม่นยำในการถอดอักษรไม่น้อยกว่าร้อยละ 90

ในส่วนของคลังข้อมูลในวิทยานิพนธ์ฉบับนี้คือคำภาษาอังกฤษและคำทับศัพท์ที่เขียนเป็นภาษาไทยโดยราชบัณฑิตยสถานจำนวน 10,060 คู่คำ เมื่อได้คลังข้อมูลที่เพียงพอแล้วจะนำคลังข้อมูลมาคัดแยกคำที่เป็นคำยืมภาษาต่างประเทศและคำที่มาจากหนังสือที่ตีพิมพ์ก่อน พ.ศ. 2532 ซึ่งเป็นปีที่ราชบัณฑิตยสถานตีพิมพ์หลักเกณฑ์การทับศัพท์ภาษาอังกฤษฉบับล่าสุดทิ้งไป

เพื่อจะทำให้กฎการถอดอักษรที่ได้ออกมาแม่นยำยิ่งขึ้น หลังจากนั้นจะนำข้อมูลมาตัดอักษรภาษาอังกฤษและภาษาไทย แล้วนำมาจับคู่ข้อมูลโดยอัตโนมัติ โดยที่อักษรภาษาอังกฤษที่อยู่ตำแหน่งเดียวกับอักษรภาษาไทยจะนำมาจับคู่กัน จากนั้นผู้วิจัยจะแก้ไขจากจับคู่ข้อมูลด้วยมืออีกครั้งเพื่อความถูกต้องของการจับคู่ข้อมูลที่ได้

หลังจากเตรียมข้อมูลเรียบร้อยแล้วก็จะเขียนโปรแกรมนำข้อมูลที่มีการจับคู่ข้อมูลเรียบร้อยแล้วนั้นแยกลงในตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 1 ตำแหน่ง 2 ตำแหน่ง และ 3 ตำแหน่งตามลำดับ หากพบการถอดอักษรที่บริบทเดียวกันสามารถถอดอักษรได้หลายแบบจะนำมาหาค่าความถี่ โดยแบบการถอดอักษรที่มีค่าความถี่มากที่สุดจะนำมาใส่ในตารางคาดเดาจะแบ่งเป็นตารางคาดเดาที่มีบริบทซ้ายและขวาอย่างและ 1 และ 2 ตำแหน่ง และตารางคาดเดาที่ไม่อาศัยบริบทเลย

เมื่อได้ตารางกฎและตารางคาดเดาเรียบร้อยแล้วจะนำกฎสำหรับคอมพิวเตอร์ที่ได้มาเปรียบเทียบความสอดคล้องกับหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน โดยหาร้อยละของความสอดคล้องของอักษรแต่ละตัว เพื่อที่จะดูว่าคลังข้อมูลของราชบัณฑิตยสถานนั้นจะสร้างกฎที่มีความสอดคล้องกับกฎการถอดอักษรของราชบัณฑิตยสถานเพียงไร ซึ่งหากเป็นไปตามสมมติฐานแล้วนั้น กฎสำหรับคอมพิวเตอร์ที่ได้มาจากคลังข้อมูลของราชบัณฑิตยสถานจะต้องสอดคล้องกับหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน ผลการเปรียบเทียบความสอดคล้องของกฎของคอมพิวเตอร์กับกฎของราชบัณฑิตยสถานปรากฏว่าการถอดสระมีความสอดคล้องร้อยละ 92.68 การถอดพยัญชนะต้นมีความสอดคล้องร้อยละ 97.35 และการถอดพยัญชนะท้ายมีความถูกต้องร้อยละ 97.34 จะเห็นได้ว่ากฎทั้งสองมีความสอดคล้องกันมาก แต่ก็ยังมีบางส่วนที่ไม่สอดคล้องกันเนื่องมาด้วยสาเหตุดังที่กล่าวไปแล้วในบทที่ 4 ซึ่งหากได้แก้ไขเพิ่มเติมก็จะทำให้การถอดอักษรครอบคลุมและมีมาตรฐานดีขึ้น

หลังจากได้ตารางกฎและตารางคาดเดาแล้ว จากนั้นจะนำตารางกฎและตารางคาดเดาที่สร้างมาจากคลังข้อมูลของราชบัณฑิตยสถานนั้นมาเป็นกฎในการถอดอักษรให้กับโปรแกรมการถอดอักษร โดยที่เมื่อโปรแกรมเจอข้อมูลที่รับเข้ามาตรงกับกฎในตารางใดก็จะให้ผลลัพธ์ตามนั้นออกมา ลำดับในการเทียบกฎนั้นโปรแกรมจะเทียบจากตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 1 ตำแหน่ง ก่อนแล้วจึงไปเทียบจากตารางกฎที่มีบริบทซ้ายและขวาอย่างละ 2 และ 3 ตำแหน่งตามลำดับ หากไม่พบจะไปเทียบจากตารางคาดเดาที่มีบริบทซ้ายและขวาอย่างละ 2 ตำแหน่ง

ก่อนแล้วจึงไปเทียบกับตารางคาดเดาที่มีบริบทซ้ายและขวาอย่างละ 1 ตำแหน่ง หากยังไม่พบจะไปเทียบกับตารางคาดเดาที่ไม่ใช้บริบทซึ่งดูจากแบบการถอดอักษรที่ปรากฏมากที่สุดในคลังข้อมูล ด้วยเหตุนี้คำที่ป้อนเข้ามาทุกคำจะสามารถถอดอักษรได้ครบถ้วน

ในการทดสอบประสิทธิภาพของโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมา ผู้วิจัยได้นำโปรแกรมการถอดอักษรไปทดลองถอดอักษรกับชุดข้อมูลทดสอบที่จัดเตรียมไว้ แล้วทำการประเมินผลประสิทธิภาพความถูกต้องของการถอดอักษรว่าเหมือนกับที่ราชบัณฑิตยสถานถอดอักษรไว้เพียงไร ผลการทดลองปรากฏว่าโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่พัฒนาขึ้นมีประสิทธิภาพในการถอดอักษรไม่สูงนัก โดยสามารถถอดอักษรได้ถูกต้องในระดับคำร้อยละ 39.53 และถอดอักษรได้ถูกต้องในระดับอักษรร้อยละ 79.8 ซึ่งไม่เป็นไปตามความคาดหมาย อาจมีสาเหตุของการถอดอักษรผิดพลาดดังที่ได้นำเสนอในบทที่ 4

ดังนั้นผู้วิจัยจึงเห็นว่าหากต้องการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยให้มีประสิทธิภาพสูงขึ้นนั้น จะอาศัยการถอดอักษรจากภาษาอังกฤษเป็นภาษาไทยโดยตรงเลยไม่ได้ ควรจะต้องมีการเปลี่ยนหรือเสริมขั้นตอนในการดำเนินการวิจัยเพื่อให้ได้โปรแกรมที่มีประสิทธิภาพมากยิ่งขึ้น ดังจะกล่าวในข้อเสนอแนะต่อไป

5.2 ข้อเสนอแนะในการศึกษาพัฒนาเพิ่มเติม

จากการที่ได้พัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานตามขั้นตอนการวิจัยดังที่ได้กล่าวไปแล้วข้างต้นนั้น พบว่าประสิทธิภาพของโปรแกรมการถอดอักษรที่พัฒนาโดยใช้วิธีเช่นวิทยานิพนธ์เล่มนี้นั้น ให้ผลความถูกต้องแม่นยำในการถอดอักษรไม่ดีเท่าที่ควรทั้งนี้เนื่องมาจากสาเหตุใหญ่ๆ 2 ด้าน คือ ด้านการถอดอักษรของข้อมูลของราชบัณฑิตยสถาน และด้านการพัฒนาโปรแกรม จึงอยากจะขอเสนอแนะแนวทางสำหรับผู้ที่ต้องการศึกษาหรือพัฒนาเพิ่มเติม โดยแบ่งข้อเสนอแนะเป็น 2 ด้าน คือ ข้อเสนอแนะด้านการถอดอักษร และข้อเสนอแนะด้านการพัฒนาโปรแกรม ดังนี้

5.2.1 ข้อเสนอแนะด้านการถอดอักษร

สำหรับด้านการถอดอักษรนั้นมีข้อสังเกตคือ กฎการถอดอักษรที่ราชบัณฑิตยสถานให้ไว้นั้นยังไม่ครอบคลุมแบบการถอดอักษรจริงทั้งหมด โดยดูได้จากการพบข้อมูลว่ามีการถอดอักษรที่นอกเหนือไปจากกฎการถอดอักษรที่ราชบัณฑิตยสถานวางไว้จำนวนหนึ่ง แบบการถอดอักษรเหล่านี้พบเป็นจำนวนมากในข้อมูลของราชบัณฑิตยสถานเองและการถอดอักษรเหล่านี้ไม่ได้มีสาเหตุมาจากความหลากหลายในการออกเสียงของผู้ถอดอักษร อีกทั้งเมื่อเทียบเสียงอ่านของคำภาษาอังกฤษแล้วพบว่าเสียงนั้นคล้ายกับอักษรไทยที่มีการถอดอักษรผิดจำนวนมากดังแสดงให้เห็นในตาราง 5-1 หากราชบัณฑิตยสถานได้แก้ไขกฎการถอดอักษรโดยเพิ่มเติมแบบการถอดอักษรที่พบนอกเหนือไปจากกฎด้วย ก็จะทำให้การการถอดอักษรของราชบัณฑิตยสถานมีความครอบคลุมการถอดอักษรมากขึ้น กฎการถอดอักษรที่น่าจะเพิ่มเติมได้มีดังนี้

สระภาษาอังกฤษ	ใช้สระภาษาไทย	จำนวนที่พบในข้อมูล	ตัวอย่าง
a (2767)	แอะ	41	bl**a**ckj**a**ck=แบล็กแจ็ก
e (2508)	#	1070	formaldehyd**e**=ฟอร์มาลดีไฮด์
e (2508)	เออ	80	coupl**e**=คัปเปิล
e (2508)	แอ	6	villan**e**lle=วิลลาแนล
er (552)	แอ	6	b**er**noulli=แบร์นูลลี
eo (48)	ออ	7	georgia=จอร์เจีย
o (1998)	อ็อ	38	**o**bstruent=อ็อบสตรูอันต์
ue (20)	#	11	analog**ue**=แอนะล็อก
yr (10)	อิ	5	p**yr**rhotite=พิร์โรไทต์

ตาราง 5-1 กฎการถอดสระเพิ่มเติมจากกฎของราชบัณฑิตยสถาน

จากตาราง 5-1 จะเห็นได้ว่า a นั้น ถอดอักษรเป็น แอะ จำนวน 41 ครั้ง จากจำนวนทั้งหมด 2767 ครั้ง e ที่ลงท้ายคำจะไม่ออกเสียงจำนวน 1,070 ครั้ง จากจำนวนทั้งหมด 2508 ครั้ง

ส่วน e ที่ตามหลัง l จะถอดอักษรเป็น เออ 80 ครั้ง สำหรับ e ที่มี lle ตามมาจะถอดอักษรเป็น แอ 6 ครั้ง เป็นต้น

พยัญชนะภาษาอังกฤษ	ใช้พยัญชนะภาษาไทย	จำนวนที่พบในข้อมูล	ตัวอย่าง
cl(735)	ก	6	wy**cl**iffe=วิกลิฟฟ์
t (1577)	จ	9	s**t**uart=สจวต
s (425)	#	29	nabbe**s**=แนบบ์

ตาราง 5-2 กฎการถอดพยัญชนะเพิ่มเติมจากกฎของราชบัณฑิตยสถาน

จากตาราง 5-2 จะเห็นได้ว่า cl ถอดอักษรเป็น ก จำนวน 6 จากจำนวนข้อมูลทั้งหมด 735 อักษร ซึ่งแสดงให้เห็นว่าในการถอดอักษรจริงนั้นนิยมถอดอักษร c ที่ตามด้วย l เป็น ก จำนวนมาก นอกจากนี้พบจากถอดอักษร t เป็น จ 9 ครั้งจาก 1577 ครั้ง และในส่วนของ s ท้ายคำนั้นโดยทั่วไปจะไม่ถอดอักษรถึง 29 ครั้งจาก 425 ครั้ง เมื่อการถอดอักษรทั่วไปเป็นเช่นนี้ ราชบัณฑิตยสถานอาจจะพิจารณากฎใหม่หรือเพิ่มเติมกฎให้ชัดเจนมากยิ่งขึ้นเช่นในกรณีของ e ท้ายคำ และ s ท้ายคำที่ดูเหมือนว่าคนทั่วไปจะเข้าใจว่าไม่ต้องถอดอักษรอยู่แล้วนั้น หากราชบัณฑิตยสถานได้เพิ่มกฎอธิบายให้ชัดเจนว่า e และ s ท้ายคำไม่ต้องถอดเป็นอักษรใดๆ ก็จะทำให้หลักเกณฑ์การทับศัพท์ครบถ้วนมากขึ้น

นอกจากข้อเสนอแนะด้านการถอดอักษรแล้วในข้อถัดไปจะเป็นข้อเสนอแนะด้านการพัฒนาโปรแกรม

5.2.2 ข้อเสนอแนะด้านการพัฒนาโปรแกรม

จากผลความถูกต้องแม่นยำของโปรแกรมที่ค่อนข้างต่ำนั้น ผู้วิจัยสันนิษฐานว่าอาจเนื่องมาจากข้อมูลที่ใช้กับโปรแกรมอาจเป็นคำที่ไม่เป็นไปตามเกณฑ์อย่างสม่ำเสมอ เพราะอาจเป็นคำเก่าที่ใช้จนติด จึงได้นำคำภาษาอังกฤษมาถอดอักษรเองตามหลักเกณฑ์การทับศัพท์ของราชบัณฑิตยสถาน เพื่อจะดูว่าโปรแกรมจะทำงานดีขึ้นไหม ทั้งนี้ผู้วิจัยคาดว่าผลการทำงานของโปรแกรมน่าจะดีขึ้นเพราะเกณฑ์ที่ราชบัณฑิตยสถานกำหนดไว้ก็เพื่อให้ใช้ทับศัพท์คำใหม่ๆ

ปรากฏว่าผลการทำงานของโปรแกรมดีขึ้นเล็กน้อย แสดงให้เห็นว่าว่าปัญหาส่วนหนึ่งมาจากข้อมูลที่ทดสอบจริง แต่ผลก็ยังคงไม่น่าพอใจ จึงต้องมีการปรับปรุงโปรแกรมต่อไป ผลการถอดอักษรคำยืมใหม่แสดงให้เห็นในตาราง 5-3

ทดสอบในระดับ	จำนวนทั้งหมด	จำนวนถอดอักษรถูกต้อง	จำนวนที่ถอดอักษรผิดพลาด	ร้อยละของความถูกต้องแม่นยำ (%)
คำ	100	55	45	55
อักษร	652	526	126	81

ตาราง 5-3 ผลการถอดอักษรคำยืมใหม่

จะเห็นได้ว่าจากตาราง 5-3 เมื่อทดสอบกับคำยืมใหม่แล้วจะได้ผลความแม่นยำของโปรแกรมในระดับคำมากกว่าคลังคำทับศัพท์ของราชบัณฑิตยสถานอยู่ร้อยละ 15.47 และในระดับอักษรมีความแม่นยำมากกว่าร้อยละ 1.2 หากได้ทดสอบกับคำยืมใหม่จำนวนมากขึ้น ค่าของความถูกต้องแม่นยำก็น่าจะมากขึ้นไปด้วย

จากการที่ได้พัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานตามขั้นตอนการวิจัยดังที่ได้กล่าวไปแล้วข้างต้นนั้น พบว่าประสิทธิภาพของโปรแกรมการถอดอักษรที่พัฒนาโดยใช้วิธีเช่นวิทยานิพนธ์เล่มนี้นั้น ให้ผลความถูกต้องแม่นยำในการถอดอักษรไม่มากนัก จึงจะขอเสนอแนะแนวทางในการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยเพื่อการศึกษาพัฒนาเพิ่มเติม ดังนี้

5.2.2.1 ควรพัฒนาโปรแกรมการถอดอักษรโดยผ่านตัวกลางเป็นรูปแทนเสียง

เมื่อได้พัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโยใช้คลังข้อมูลของราชบัณฑิตยสถานไปแล้วนั้น ผู้วิจัยเลือกที่จะแปลงข้อมูลภาษาอังกฤษเป็นภาษาไทยโดยตรงเพราะคิดว่าคลังข้อมูลน่าจะสร้างกฎการถอดอักษรที่ถูกต้องแม่นยำออกมาได้ แต่จากผลการ

ทดลองถอดอักษรทำให้เห็นว่าการแปลงสัญลักษณ์จากรูปอักษรอังกฤษไปเป็นรูปอักษรไทยโดยตรงนั้นจะมีความถูกต้องในการถอดอักษรไม่มากเท่าที่ควร ผู้วิจัยจะขอเสนอแนะให้ทดลองพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยมีการแปลงรูปอักษรภาษาอังกฤษเป็นหน่วยเสียงของภาษาอังกฤษก่อน แล้วจึงไปเทียบกับหน่วยเสียงภาษาไทย แล้วจึงแปลงเป็นรูปอักษรภาษาไทยต่อไป ดังเช่น งานของ Knight and Graehl (1997) ซึ่งศึกษาเรื่อง การถอดอักษรคาตาคานะในภาษาญี่ปุ่นที่เขียนทับศัพท์ภาษาอังกฤษกลับไปเป็นภาษาอังกฤษโดยใช้เครื่องคอมพิวเตอร์ ได้ใช้แนวทางการถอดอักษรโดยใช้ตัวกลาง เพราะในการถอดอักษรญี่ปุ่นที่เขียนทับศัพท์กลับไปเป็นภาษาอังกฤษนั้นจำเป็นต้องใช้ตัวกลางคือรูปแทนเสียง แล้วจึงถอดกลับไปเป็นอักษรภาษาอังกฤษได้ ผู้วิจัยเห็นว่าการพัฒนาโปรแกรมการถอดอักษรโดยใช้รูปแทนเสียงจะให้ผลของข้อมูลที่ถูกต้องแม่นยำกว่าการถอดอักษรโดยไม่ใช้ตัวกลางเพราะการถอดอักษรจะได้ถูกต้องตามการออกเสียงภาษาอังกฤษของคำที่ต้องการจะถอด นอกจากนี้การใช้รูปแทนเสียงจะกำจัดปัญหาความหลากหลายในการถอดอักษรเพราะผู้ถอดอักษรออกเสียงคำภาษาอังกฤษแตกต่างกันได้ด้วย แต่ในขณะเดียวก็ต้องพิจารณาปัญหาความถูกต้องของเสียงที่ได้ออกมาด้วย เพราะเมื่อแปลงคำภาษาอังกฤษ เป็นรูปแทนเสียงภาษาอังกฤษ และถอดอักษรออกมานั้นก็อาจจะไม่ได้ผลความถูกต้องร้อยละ 100 ยกตัวอย่างเช่นงานของ Knight and Graehl (1997) ที่กล่าวไปข้างต้นนั้นได้ผลของความถูกต้องร้อยละ 64 ซึ่งก็แสดงให้เห็นว่าแม้ใช้รูปแทนเสียงแล้วก็ยังสามารถเกิดข้อผิดพลาดขึ้นได้

5.2.2.2 **ควรใช้คลังข้อมูลที่มีขนาดใหญ่ขึ้นและมีคุณภาพมากขึ้น**

หากต้องการพัฒนาโปรแกรมการถอดอักษรโดยใช้คลังข้อมูลนั้น เพื่อให้ได้ประสิทธิภาพมากขึ้น ควรใช้คลังข้อมูลฝึกสอนที่มีขนาดใหญ่ขึ้นเพื่อให้โปรแกรมได้เรียนรู้แบบการถอดอักษรต่างๆ และนำมาคำนวณค่าความน่าจะเป็นทางสถิติได้ดียิ่งขึ้น นอกจากนี้การเพิ่มขนาดของข้อมูลนั้นควรเพิ่มทั้งในเชิงปริมาณและคุณภาพ กล่าวคือ มีค่าความถี่ในการปรากฏของแบบการถอดอักษรต่างๆมากขึ้น รวมทั้งข้อมูลการถอดอักษรที่เก็บรวบรวมมาต้องมีการถอดอักษรที่ถูกต้องตามหลักเกณฑ์ มีการจัดเก็บและจับคู่คำที่ถูกต้อง นอกจากนี้แบบการถอดอักษรที่ได้ควรจะมีมาตรฐานที่ถูกต้องเพียงแบบเดียวเพื่อลดความผิดพลาดในการถอดอักษรนั่นเอง

5.2.2.3 **ควรกำหนดขอบเขตพยางค์ให้กับคลังข้อมูล**

จะเห็นได้จากในบทที่ 4 ว่าคลังข้อมูลภาษาที่ไม่มีการกำหนดขอบเขตพยางค์ให้กับคลังข้อมูลนั้น จะทำให้เกิดความผิดพลาดในการแยกพยางค์และการกำหนดพยัญชนะต้นและพยัญชนะท้ายได้ดังที่ได้กล่าวไปแล้ว ดังนั้นคลังข้อมูลภาษาที่เก็บรวบรวมมาควรมีการกำหนดขอบเขตพยางค์ให้เรียบร้อย เพื่อที่จะได้แยกแยะพยัญชนะต้นหรือพยัญชนะท้ายในพยางค์ได้ และจะทำให้การถอดอักษรมีประสิทธิภาพมากยิ่งขึ้น

หลังจากได้เห็นกระบวนการการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยโดยใช้คลังคำทับศัพท์ของราชบัณฑิตยสถานและผลการทดลองถอดอักษรของโปรแกรมที่พัฒนาขึ้นมาแล้วนั้น ถ้ามีผู้สนใจนำวิธีการที่เสนอแนะไปประยุกต์ใช้ในการพัฒนาโปรแกรมการถอดอักษรภาษาอังกฤษเป็นภาษาไทยที่มีประสิทธิภาพมากขึ้นต่อไป ก็จะทำให้ได้โปรแกรมการถอดอักษรที่มีความถูกต้องแม่นยำ และหากคนทั่วไปได้ใช้โปรแกรมการถอดอักษรตามหลักราชบัณฑิตยสถานในการถอดอักษรภาษาอังกฤษเป็นภาษาไทย ก็จะทำให้การถอดอักษรเป็นไปอย่างมีระบบและเป็นมาตรฐานเดียวกันมากยิ่งขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

กาญจนา นาคสกุล. (ม.ป.ป.). การเขียนคำทับศัพท์ภาษาอังกฤษ. (ม.ป.ท.).

ชนินทร์ชัย อินทิราภรณ์. 2543. พจนานุกรมคำทับศัพท์ ภาษาอังกฤษ ฝรั่งเศส เยอรมัน อิตาลี สเปน ญี่ปุ่น รัสเซีย อาหรับและมลายู. กรุงเทพมหานคร: ไอ.คิว.บุ๊คเซ็นเตอร์.

ด้วง พุธศุกร์. 2544. ศัพท์วิทยาศาสตร์และศัพท์วิชาการ อังกฤษ-ไทย. เชียงใหม่: หน่วยงานเทคโนโลยีการพิมพ์และสารสนเทศ มหาวิทยาลัยเชียงใหม่.

ทัศนวรรณ ศูนย์กลาง สมชาย ประสิทธิ์จูตระกูล และบุญเสริม กิจศิริกุล 2543. (735)การเข้ารหัสคำทับศัพท์ภาษาไทย/อังกฤษเพื่อการค้นคืนข้ามภาษาด้วยเทคนิคนิวรอลเน็ตเวิร์ก หนังสือรวมเอกสารวิชาการของการประชุมวิชาการวิทยาศาสตร์และวิศวกรรมคอมพิวเตอร์แห่งชาติ ครั้งที่ 4 : 158-165.

ธนิส อินทิราภรณ์. 2543. พจนานุกรมศัพท์เทคนิค. กรุงเทพมหานคร: ไอ.คิว.บุ๊คเซ็นเตอร์.

บุญเสริม ฤทธาธิรมย์. 2522. อธิบายคำทับศัพท์ภาษาอังกฤษ. กรุงเทพ: โรงพิมพ์รุ่งเรืองรัตน์.

ฝ่ายวิชาการสำนักพิมพ์. 2546. พจนานุกรมคำทับศัพท์ภาษาอังกฤษ. กรุงเทพมหานคร: สำนักพิมพ์วาดศิลป์.

พรทิพย์ แฟงสุด. 2542. พจนานุกรมคำทับศัพท์ภาษาต่างประเทศ. กรุงเทพมหานคร: สำนักพิมพ์ฟิสิกส์เซ็นเตอร์.

พิณทิพย์ ทวยเจริญ. 2544. ภาพรวมของการศึกษาภาษาและภาษาศาสตร์. กรุงเทพมหานคร: สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์.

ราชบัณฑิตยสถาน. 2547. สารานุกรมประวัติศาสตร์สากล: อเมริกา เล่ม 1 อักษร A-B. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2546. ศัพท์วิทยาศาสตร์. พิมพ์ครั้งที่ 5. กรุงเทพมหานคร: อรุณการพิมพ์

ราชบัณฑิตยสถาน. 2546. ศัพท์ภาษาศาสตร์. นนทบุรี: สหมิตรพรินติ้ง.

ราชบัณฑิตยสถาน. 2545. ศัพท์คณิตศาสตร์. พิมพ์ครั้งที่ 8. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2545. พจนานุกรมศัพท์วรรณกรรม อังกฤษ-ไทย. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2544. พจนานุกรมศัพท์ธรณีวิทยา. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2542. ศัพท์เทคโนโลยีสารสนเทศฉบับราชบัณฑิตยสถาน. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2542. ศัพท์การเชื่อม. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2542. พจนานุกรมศัพท์ยานยนต์. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2541. ศัพท์พลังงาน อังกฤษ-ไทย. กรุงเทพมหานคร: อรุณการพิมพ์

ราชบัณฑิตยสถาน. 2541. พจนานุกรมศัพท์ปรับอากาศ. กรุงเทพมหานคร: อรุณการพิมพ์.

ราชบัณฑิตยสถาน. 2539. สารานุกรมประวัติศาสตร์สากลสมัยใหม่: เอเชีย เล่ม 1 อักษร A-B ฉบับราชบัณฑิตยสถาน. กรุงเทพมหานคร: โรงพิมพ์มหาจุฬาลงกรณราชวิทยาลัย.

ราชบัณฑิตยสถาน. 2537. สารานุกรมประวัติศาสตร์สากลสมัยใหม่: ยุโรป เล่ม 2 อักษร C-D ฉบับราชบัณฑิตยสถาน. กรุงเทพมหานคร: มหาจุฬาลงกรณราชวิทยาลัย.

ราชบัณฑิตยสถาน. 2537. พจนานุกรมศัพท์สัทศาสตร์. กรุงเทพมหานคร: เพื่อนพิมพ์.

ราชบัณฑิตยสถาน. 2532. หลักเกณฑ์การทับศัพท์ภาษาอังกฤษ. กรุงเทพมหานคร: คุรุสภา.

ราชบัณฑิตยสถาน. 2523. พจนานุกรมศัพท์ภูมิศาสตร์ อังกฤษ-ไทย ฉบับราชบัณฑิตยสถาน เล่ม 1 A-K. กรุงเทพมหานคร: นนทชัย.

ราชบัณฑิตยสถาน. 2523. พจนานุกรมศัพท์ภูมิศาสตร์ อังกฤษ-ไทย ฉบับราชบัณฑิตยสถาน เล่ม 2 L-Z. กรุงเทพมหานคร: นนทชัย.

ราชบัณฑิตยสถาน. 2508. ความรู้ทางอักษรศาสตร์. พิมพ์ครั้งที่3. กรุงเทพมหานคร: นนทชัย.

วิโรจน์ อรุณมานะกุล. 2545. ภาษาศาสตร์คลังข้อมูล. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย.

สุทธิวงศ์ พงศ์ไพบูลย์. 2516. หลักภาษาไทย. กรุงเทพมหานคร: ไทยวัฒนาพานิช.

อุไรรัตน์ บุญภานนท์. 2529. การถอดอักษรอังกฤษเป็นไทยโดยใช้หลักวิชาภาษาศาสตร์ วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาภาษาศาสตร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

Catford, J.C. 1974. A Linguistic Theory of Translation. London: Oxford University Press.

Charoenporn, Thatsanee; Chotimongkol, Ananlada and Sornlertlamvanich, Virach. 1999. Automatic Romanization for Thai, Proceedings of the Second International Workshop on East-Asian Language Resources and Evaluation (ORIENTAL COCOSDA '99), pp 137-140.

Fujii, Atsushi and Ishikawa, Tetsuya. 1999. Cross-Language Information Retrieval for Technical Documents. In Proceedings of the Joint ACL SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pp.29-37.

Fujii, Atsushi and Ishikawa, Tetsuya. 2001. Japanese/English Cross-Language Information Retrieval: Exploration of Query Translation and Transliteration. In Computers and the Humanities, 35, 4: 389-420.

Gimson, A.C. and Ramsaran, S.M. 1982. An English Pronunciation Companion to the Oxford Advanced Learner's Dictionary of Current English. Oxford: Oxford University Press.

Glover, Bonnie and Knight, Kevin. 1998. Translating Names and Technical Terms in Arabic Text. COLING/ACL Workshop on Computational Approaches to Semitic Languages. Montreal, Quebéc.

Jung, SungYoung; Hong, SungLim; and Paek, Eunok. 2000. An English to Korean Transliteration Model of Extended Markov Window. In Proceedings of the 18th International Conference on Computational Linguistics (COLING00), volume 1, pages 383-389, Saarbrüken.

Kang, In-Ho and Kim, GilChang. 2000. English-to-Korean Transliteration using Multiple Unbounded Overlapping Phoneme Chunks. In Proceedings of the 18th International Conference on Computational Linguistics (COLING00), volume 1, pages 418-424, Saarbrüken.

Kawtrakul, Asanee; Deemagarn, Amarin; Thumkanon, Chalathip; and Khantonthong, Navapat. 1998. Backward Transliteration for Thai Document Retrieval. IEEE Asia Pacific Conference on Circuits and Systems.

Knight, Kevin and Graehl, Jonathan. 1997. Machine Transliteration. Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics.

Leonard Bloomfield. 1974. Language. New York: Holt, Rinehart and Winston.

Suwanvisat, Prayut and Prasitjutrakul, Somchai. 1998. Thai-English Cross-Language Transliterated Word Retrieval using Soundex Technique. the National Computer Science and Engineering Conference, Kasetsart University, Bangkok, Thailand.

Suwanvisat, Prayut and Prasitjutrakul, Somchai. 1999. Transliterated Word Encoding and Retrieval Algorithms for Thai-English Cross-Language Retrieval. the National Computer Science and Engineering Conference, Bangkok, Thailand.

Van den Bosch, Antal and Daelemans, Walter. 1993. Data-Oriented Methods for Grapheme-to-Phoneme Conversion. Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics, Utrecht, The Netherlands.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ตารางที่ ก-1: เปรียบเทียบความสอดคล้องของกฎการถอดสระภาษาอังกฤษเป็นภาษาไทยของคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		A (2767)		
a	แอ	b**a**dminton = แบดมินตัน	835	31.10
a	อะ	**a**luminium =อะลูมิเนียม	395	14.71
a	อา	Chic**a**go = ชิคาโก	1033	38.47
a	เอ	**A**sia = เอเชีย	294	10.95
a	ออ	footb**a**ll=ฟุตบอล	72	2.68
a สอดคล้อง			2704	97.72
a	อาร์	b**a**zett=บาร์เซตต์	1	0.04
a	อิ	anchor**a**ge=แองคอริจ	1	0.04
a	อี	macaul**a**y=แมกคอลีย์	2	0.07
a	เอะ	templ**a**te=เทมเพล็ต	1	0.04
a	เออ	sew**a**rd=ซีเวิร์ด	3	0.11
a	แอะ	bl**a**ckjack=แบล็กแจ็ก	41	1.53
a	#	vaugh**a**n=วอน	6	0.22
a	โอ	m**a**noid=โมนอยด์	1	0.04
a ไม่สอดคล้อง			63	2.28
		AR (455)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ar	แอ	**ar**row=แอร์โรว์	39	8.57
ar	อา	b**ar**=บาร์	408	89.67
ar	ออ	w**ar**d=วอร์ด	5	1.10
ar	เออ	Edw**ar**d=เอดเวิร์ด	3	0.66
ar สอดคล้อง			455	100.00
ar ไม่สอดคล้อง			0	0.00
		ARE (24)		
are	แอ	m**are**=แมร์	14	58.33
are	อา	**are**=อาร์	10	41.67
are สอดคล้อง			24	100.00
areไม่สอดคล้อง			0	0.00
		AA (9)		
aa	อา	baz**aa**r=บาซาร์	4	44.44
aa	แอ	**Aa**ron=แอรอน	4	44.44
aa สอดคล้อง			8	
aa	อะ	k**aa**pprovinsie=คัปโพรวินซี	1	11.11
aa ไม่สอดคล้อง			1	11.11
		AE (18)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ae	อี	**Ae**gean=อีเจียน	8	44.44
ae	แอ	**ae**rosphere=แอโรสเฟียร์	1	5.56
ae	เอ	sund**ae**=ซันเด	8	44.44
ae สอดคล้อง			17	94.44
ae	เออ	mich**ae**l=ไมเคิล	1	5.56
ae ไม่สอดคล้อง			1	5.56
		AEA (2)		
aea	เอีย	Jud**aea**=จูเดีย	2	100.00
aea สอดคล้อง			2	100.00
aeaไม่สอดคล้อง			0	0.00
		AER (1)		
aer	เออ	k**aer**sutite=เคอร์ซูไทต์	1	100.00
aer สอดคล้อง			1	100.00
aerไม่สอดคล้อง			0	0.00
		AI (66)		
ai	เอ	Sp**ai**n=สเปน	46	69.70
ai	ไอ	C**ai**ro=ไคโร	19	28.79

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ai สอดคล้อง			65	98.48
ai	อาย	haw**ai**i=ฮาวาย	1	1.52
aiไม่สอดคล้อง			1	1.52
		AIR (5)		
air	แอ	Bel **Air**=เบลแอร์	5	100.00
air สอดคล้อง			5	100.00
airไม่สอดคล้อง			0	0.00
		AO (3)		
ao	เอา	Mindan**ao**=มินดาเนา	2	66.67
ao สอดคล้อง			2	66.67
ao	อาว	l**ao**s=ลาว	1	33.33
aoไม่สอดคล้อง			1	33.33
		AU (78)		
au	อา	l**au**gh=ลาฟ	3	3.85
au	ออ	**Au**gusta=ออกัสตา	61	78.21
au	เอา	Biss**au**=บิสเซา	9	11.54
au	โอ	**Au**ger=โอเจอร์	3	3.85
au	แอ	L**au**ghlin=แลฟลิน	1	1.28

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
au สอดคล้อง			77	98.72
au	เอาะ	b**au**xite=บ็อกไซต์	1	1.28
auไม่สอดคล้อง			1	1.28
		AW (18)		
aw	ออ	l**aw**ernce=ลอว์เรนซ์	18	100.00
aw สอดคล้อง			18	100.00
aw ไม่สอดคล้อง			0	0.00
		AY (48)		
ay	เอ	Mal**ay**=มาเลย์	44	91.67
ay สอดคล้อง			44	91.67
ay	อี	barcl**ay**=บาร์คลีย์	1	2.08
ay	ไอ	um**ay**yah=อุไมยะห์	3	6.25
ayไม่สอดคล้อง			4	8.33
		AYR (1)		
ayr	แอ	**Ayr**shire=แอร์เชอร์	1	100.00
ayr สอดคล้อง			1	100.00

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ayrไม่สอดคล้อง			0	0.00
		E (2508)		
e	อี	Sweden=สวีเดน	239	9.53
e	เอ	Lebanon=เลบานอน	953	38.00
e	อิ	electronics =อิเล็กทรอนิกส์	87	3.47
e	เอะ	Mexico =เม็กซิโก	70	2.79
e สอดคล้อง			1349	
e	#	formaldehyde=ฟอร์มาลดีไฮด์	1070	42.66
e	ออ	becquerel=เบ็กเคอรอล	1	0.04
e	อา	catene=แคทีนา	1	0.04
e	เออ	couple=คัปเปิล	80	3.19
e	เออะ	libreville=ลิเบรอะวิล	1	0.04
e	แอ	villanelle=วิลลาแนล	6	0.24
eไม่สอดคล้อง			1159	46.21
		ER (552)		
er	เออ	Canberra=แคนเบอร์รา	544	98.55
er	อา	Clerk = คลาร์ก	2	0.36
er สอดคล้อง			546	
er	แอ	bernoulli=แบร์นูลลี	6	1.09

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
erไม่สอดคล้อง			6	1.09
		ERE (11)		
ere	เอีย	cashm**ere** = แคชเมียร์	10	90.91
ere สอดคล้อง			10	
ere	แอ	amp**ere**=แอมแปร์	1	9.09
ereไม่สอดคล้อง			1	9.09
		EA (77)		
ea	อี	Guin**ea**=กินี	45	58.44
ea	เอ	D**ea**d Sea=เดดซี	20	25.97
ea	เอีย	Caribb**ea**n=แคริบเบียน	11	14.29
ea สอดคล้อง			76	
ea	อา	gr**ea**ssmann=กราสมันน์	1	1.30
eaไม่สอดคล้อง			1	1.30
		EAR (15)		
ear	แอ	B**ear**= แบร์	2	13.33
ear	เอีย	g**ear**= เกียร์	9	60.00
ear	อา	H**ear**t =ฮาร์ต	1	6.67
ear	เออ	P**ear**l Harbour =เพิร์ลฮาร์เบอร์	3	20.00

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ear สอดคล้อง			15	100.00
earไม่สอดคล้อง			0	0.00
		EAU (5)		
eau	โอ	B**eau**fort =โบฟอร์ด	4	80.00
eau	อิว	B**eau**ty = บิวตี	1	20.00
eau สอดคล้อง			5	100.00
eauไม่สอดคล้อง			0	0.00
		EE (56)		
ee	อี	Gr**ee**nwich = กรีนิช	56	100.00
ee สอดคล้อง			56	100.00
eeไม่สอดคล้อง			0	0.00
		EER (4)		
eer	เอีย	b**eer** = เบียร์	4	100.00
eer สอดคล้อง			4	100.00
eerไม่สอดคล้อง			0	0.00
		EI (40)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ei	อี	N**ei**l= นีล	10	25.00
ei	ไอ	**Ei**nsteinium = ไอน์สไตเนียม	23	57.50
ei	เอ	B**ei**rut =เบรุต	7	17.50
ei สอดคล้อง			40	
eiไม่สอดคล้อง			0	0.00
		EIR (4)		
eir	แอ	h**eir** = แอร์	1	25.00
eir	เอีย	P**eir**se= เพียร์ส	2	50.00
eir	เออ	p**eir**ce=เพิร์ซ	1	25.00
eir สอดคล้อง			4	
eirไม่สอดคล้อง			0	0.00
		EO (48)		
eo	อี	p**eo**ple = พีเพิล	30	62.50
eo	เอ	L**eo**minster = เลมินสเตอร์	8	16.67
eo	เอีย	Napol**eo**n = นะโปเลียน	1	2.08
eo	เอียว	Born**eo** =บอร์เนียว	2	4.17
eo สอดคล้อง			41	
eo	ออ	g**eo**rgia=จอร์เจีย	7	14.58

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
eoไม่สอดคล้อง			7	14.58
		EOU (3)		
eou	โอ	S**eou**l =โซล	1	33.33
eou สอดคล้อง			1	
eou	เอีย	silic**eou**s=ซีลิเชียส	2	66.67
eouไม่สอดคล้อง			2	66.67
		EU (34)		
eu	อิว	l**eu**kemia =ลิวคีเมีย	14	41.18
eu	ยุ	**Eu**rope=ยุโรป	1	2.94
eu	ยู	**Eu**phrates =ยูเฟรทีส	8	23.53
eu	เอีย	ol**eu**m= โอเลียม	4	11.76
eu	อู	R**eu**ben= รูเบ็น	4	11.76
eu สอดคล้อง			31	
eu	ออย	**eu**ther=รอยเทอร์	1	2.94
eu	เออ	tourn**eur**=เทอร์เนอร์	2	5.88
euไม่สอดคล้อง			3	8.82
		EUR (2)		
eur	เออ	f**leur**-de-lis=เฟลอร์เดอลีส์	2	100.00

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
eur สอดคล้อง			2	
eurไม่สอดคล้อง			0	0.00
		EW (26)		
ew	อิว	New York = นิวยอร์ก	17	65.38
ew	โอ	sew = โซว์	1	3.85
ew	อู	Andrew = แอนดรูว์	8	30.77
ew สอดคล้อง			26	
ewไม่สอดคล้อง			0	0.00
		EY (85)		
ey	เอ	Yardley = ยาร์ดเลย์	19	22.35
ey	อี	key = คีย์	63	74.12
ey สอดคล้อง			82	
ey	ไอ	feynman=ไฟยน์แมน	3	3.53
eyไม่สอดคล้อง			3	3.53
		I (2205)		
i	อิ	King = คิง	1481	67.17
i	อี	ski = สกี	127	5.76
i	ไอ	Liberia = ไลบีเรีย	592	26.85

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
i สอดคล้อง			2200	
i	#	lewiston=ลุยสตัน	1	0.05
i	อะ	wilbye=วัลบี	1	0.05
i	อาย	phi=ฟาย	3	0.14
iไม่สอดคล้อง			5	0.23
		IR (32)		
ir	เออ	zircon = เซอร์คอน	24	75.00
ir	อี	Pamir = ปามีร์	6	18.75
ir สอดคล้อง			30	
ir	เอีย	sapir=ซาเพียร์	2	6.25
ir ไม่สอดคล้อง			2	6.25
		IRE (19)		
ire	ไอ	Ireland = ไอร์แลนด์	7	36.84
ire	เออ	Hampshire=แฮมป์เชอร์	1	5.26
ire	เอีย	Hampshire= แฮมป์เชียร์	11	57.89
ire สอดคล้อง			19	
ireไม่สอดคล้อง			0	0.00
		IA (194)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ia	เอีย	Ind**ia** = อินเดีย	194	100.00
ia สอดคล้อง			194	
iaไม่สอดคล้อง			0	0.00
		IE (72)		
ie	อี	r**ie**beckite= รีเบกไกต์	51	70.83
ie	เอีย	Sovi**e**t = โซเวียต	11	15.28
ie	อาย	p**ie**=พาย	2	2.78
ie	ไอ	neckt**ie**= เนกไท	7	9.72
ie สอดคล้อง			71	
ie	เอ	ern**ie**=เออร์เน	1	1.39
ieไม่สอดคล้อง		81	1	1.39
		IER (13)		
ier	เอีย	glac**ier**=เกลเชียร์	13	100.00
ier สอดคล้อง			13	
ierไม่สอดคล้อง			0	0.00
		IEW (1)		
iew	อิว	v**iew** =วิว	1	100.00

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
iew สอดคล้อง			1	100.00
iewไม่สอดคล้อง			0	0.00
		ION (47)		
ion	เอียน	Un**ion**= ยูเนียน	5	10.64
ion	อัน	lot**ion**=โลชัน	41	87.23
ion สอดคล้อง			46	
ion	เอิน	on**ion**s=อันเนินส์	1	2.13
ionไม่สอดคล้อง			1	2.13
		IU (98)		
iu	เอีย	alumin**iu**m=อะลูมิเนียม	92	100.00
iu สอดคล้อง			92	
iuไม่สอดคล้อง			0	0.00
		O (1998)		
o	โอ	Cair**o**= ไคโร	1135	56.81
o	ออ	T**o**m = ทอม	669	33.48
o	อะ	Washingt**o**n = วอชิงตัน	145	7.26
o	อู	T**o**day = ทูเดย์	3	0.15

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
o สอดคล้อง			1952	
o	อุ	w**o**lfram=วุลแฟรม	3	0.15
o	อ็อ	**o**bstruent=อ็อบสตรูอันต์	38	1.90
o	อา	dr**o**vite=ดราไวต์	1	0.05
o	เออ	l**o**ve=เลิฟ	4	0.20
oไม่สอดคล้อง			46	2.30
		OR (280)		
or	ออ	c**or**ruption = คอร์รัปชัน	227	81.07
or	เออ	Winds**or**=วินด์เซอร์	51	18.21
or สอดคล้อง			278	
or	อุ	w**or**cester=วุสเตอร์	1	0.36
or	เอ	antif**or**ro=แอนติเฟร์โร	1	0.36
orไม่สอดคล้อง			2	0.71
		ORE (33)		
ore	ออ	Thomas M**ore**=ทอมัส มอร์	31	93.94
ore	โอ	Ben M**ore**=เบนโมร์	2	6.06
ore สอดคล้อง			33	100.00
oreไม่สอดคล้อง			0	0.00

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		OA (29)		
oa	ออ	Br**oa**dway = บรอดเวย์	2	6.90
oa	โอ	**Oa**kland = โอกแลนด์	24	82.76
oa	อัว	Sam**oa**=ซามัว	3	10.34
oa สอดคล้อง			29	100.00
oaไม่สอดคล้อง			0	0.00
		OAR (2)		
oar	ออ	b**oar**d = บอร์ด	2	100.00
oar สอดคล้อง			2	
oarไม่สอดคล้อง			0	0.00
		OE (41)		
oe	โอ	J**oe** =โจ	20	48.78
oe	อู	Sh**oe**maker = ชูเมกเกอร์	3	7.32
oe	เออ	g**oe**thite =เกอไทต์	6	14.64
oe	อี	Ph**oe**nix= ฟีนิกซ์	7	17.06
oe สอดคล้อง			36	
oe	อิ	ph**oe**nician=ฟินิเชียน	1	2.44
oe	อุ	s**oe**mba=ซุมบา	2	4.88
oe	เอ	r**oe**thke=เรทกี	2	4.88

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
oeไม่สอดคล้อง			5	12.20
		OER (3)		
oer	เออ	**oer**likon =เออร์ลิคอน	2	66.67
oer สอดคล้อง			2	
oer	อัว	b**oer**=บัวร์	1	33.33
oerไม่สอดคล้อง			1	33.33
		OI (57)		
oi	ออย	thyr**oi**d = ไทรอยด์	57	100.00
oi สอดคล้อง			57	
oiไม่สอดคล้อง			0	0.00
		OO (59)		
oo	อุ	f**oo**t = ฟุต	13	21.31
oo	อู	w**oo**d =วูด	47	77.05
oo	อะ	Bl**oo**dsworth= บลัดส์เวิร์ท	1	1.64
oo สอดคล้อง			61	100.00
ooไม่สอดคล้อง			0	0.00
		OOR (6)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
oor	อัว	M**oor** = มัวร์	1	16.67
oor	ออ	d**oor** =ดอร์	2	33.33
oor	โอ	D**oor**n=โดร์น	3	50.00
oor สอดคล้อง			6	100.00
oorไม่สอดคล้อง			0	0.00
		OU (81)		
ou	เอา	c**ou**nter =เคาน์เตอร์	23	28.40
ou	อาว	gr**ou**nd = กราวนด์	2	2.47
ou	โอ	B**ou**lder = โบลเดอร์	2	2.47
ou	อะ	thermoc**ou**ple= เทอร์โมคัปเปิล	18	22.22
ou	ออ	G**ou**gh = กอฟ	4	4.93
ou	อุ	s**ou**p = ซุป	9	11.11
ou	อู	Vanc**ou**ver = แวนคูเวอร์	23	28.40
ou สอดคล้อง			81	100.00
ouไม่สอดคล้อง			0	0.00
		OUR (13)		
our	ออ	b**our**nonite=บอร์โนไนต์	3	23.08
our	เออ	Melb**our**ne = เมลเบิร์น	6	46.15
our	อัว	t**our** = ทัวร์	3	23.08
our	โอ	M**our**ne= โมร์น	1	7.69

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
our สอดคล้อง			13	100.00
ourไม่สอดคล้อง			0	0.00
		OW (73)		
ow	โอ	b**ow**ling =โบว์ลิง	35	47.95
ow	เอา	C**ow**pens = เคาว์เพนส์	14	19.17
ow	อาว	t**ow**nhouse = ทาวน์เฮาส์	21	28.77
ow	อู	C**ow**per = คูว์เปอร์	1	1.37
ow สอดคล้อง			71	
ow	ออ	l**ow**ry=ลอว์รี	2	2.74
owไม่สอดคล้อง			2	2.74
		OY (21)		
oy	ออย	Ll**oy**d = ลอยด์	21	100.00
oy สอดคล้อง			21	
oyไม่สอดคล้อง			0	0.00
		U (592)		
u	อะ	H**u**ngary = ฮังการี	266	44.93
u	อิว	C**u**ba = คิวบา	53	8.95
u	อุ	Lilli**u**t = ลิลลิพุต	73	12.33

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
u	อู	Kuwait = คูเวต	178	30.07
u	ยู	Uranium = ยูเรเนียม	17	2.87
u	อิ	busy =บิซี	2	0.34
u สอดคล้อง			589	
u	ออ	nut=นอต	1	0.17
u	เออ	vanbrugh=แวนเบรอ	2	0.34
uไม่สอดคล้อง			3	0.51
		UR (62)		
ur	เออ	hurricane = เฮอร์ริเคน	62	100.00
ur สอดคล้อง			62	
urไม่สอดคล้อง			0	0.00
		URE (6)		
ure	อัว	Sure = ชัวร์	2	33.33
ure	เออ	lecture = เลกเชอร์	3	50.00
ure	เอียว	Pure = เพียวร์	1	16.67
ure สอดคล้อง			6	100.00
ureไม่สอดคล้อง			0	0.00
		UA (37)		

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ua	อัว	G**ua**dalupe= กัวดาลูป	37	100.00
ua สอดคล้อง			37	100.00
uaไม่สอดคล้อง			0	0.00
		UE (20)		
ue	อิว	T**ue**sday = ทิวส์เดย์	3	15.00
ue	อู	w**ue**stite = วูสไทต์	5	25.00
ue สอดคล้อง			8	
ue	#	analog**ue**=แอนะล็อก	11	55.00
ue	เออ	becq**ue**rel=เบ็กเคอรอล	1	5.00
ueไม่สอดคล้อง			12	60.00
		UI (15)		
ui	อุ	fr**ui**t = ฟรุต	1	6.67
ui	อู	j**ui**ce = จูซ	5	33.33
ui	อิ	circ**ui**t = เซอร์คิต	6	40.00
ui	ไอ	R**ui**slip = ไรสลิป	1	6.67
ui สอดคล้อง			13	
ui	อี	bang**ui**=บังกี	2	13.33
uiไม่สอดคล้อง			2	13.33

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		UIR (1)		
uir	อิว	Muir = มิวร์	1	100.00
uir สอดคล้อง			1	100.00
uirไม่สอดคล้อง			0	0.00
		UY (4)		
uy	อาย	Guy= กาย	3	75.00
uy สอดคล้อง			3	75.00
uy	อู	hakluyt=แฮกลูต	1	25.00
uyไม่สอดคล้อง			1	25.00
		Y (435)		
y	อิ	Odyssey = โอดิสซีย์	129	29.66
y	อี	Syria = ซีเรีย	175	40.23
y	ไอ	cyclone = ไซโคลน	127	29.20
y สอดคล้อง			431	
y	#	pepys=พีปส์	1	0.23
y	อา	mylayopolynesian=มาลาโยโพลินีเซียน	1	0.23
y	อาย	fry=ฟราย	2	0.46

สระภาษาอังกฤษ	ถอดอักษรเป็นสระภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
yไม่สอดคล้อง			4	0.92
		YE (7)		
ye	ไอ	r**ye** =ไรย์	4	57.14
ye สอดคล้อง			4	
ye	อาย	deb**ye**=เดบาย	1	14.29
ye	อี	wilb**ye**=วัลบี	1	14.29
ye	เอ	b**ye**lorussia=เบโลรุสเซีย	1	14.29
yeไม่สอดคล้อง			3	42.86
		YR (10)		
yr	เออ	M**yr**na =เมอร์นา	5	50.00
yr สอดคล้อง			5	
yr	อิ	p**yr**rhotite=พิร์โรไทต์	5	50.00
yrไม่สอดคล้อง			5	50.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

ตารางที่ ข-1: เปรียบเทียบความสอดคล้องของกฎการถอดพยัญชนะภาษาอังกฤษเป็นภาษาไทยของคอมพิวเตอร์กับกฎของราชบัณฑิตยสถาน

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		B (648)				B (12)		
b	บ	**b**ase=เบส	647	99.85	บ	Gi**bb**=กิบบ์	12	100.00
bสอดคล้อง			647	99.85			15	100.00
b	#	hob**b**es=ฮอบส์	1	0.15				
bไม่สอดคล้อง			1	0.15			0	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		C+(a,o,u,r,l) (735)				C (171)		
ca (243)	ค	**ca**t=แคต	193	26.26	ก	cubic= คิวบิก	170	99.42
ca	ก	cara**ca**s= คารากัส	45	6.12				
co (255)	ค	**co**ne=โคน	219	29.80				
co	ก	vas**co**=วัสโก	33	4.49				
cu (37)	ค	**Cu**ba= คิวบา	32	4.35				
cr (100)	ค	**cr**own= คราวน์	95	12.93				
cl (100)	ค	**Cl**eo=คลีโอ	94	12.79				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
cสอดคล้อง			711	96.74			170	99.42
ca	กค	ma**ca**ulay=แมกคอลีย์	1	0.14	กซ	hel**ic**=ฮีลิกซ์	1	0.58
ca	ซ	**ca**esium=ซีเซียม	4	0.54				
co	กก	cho**co**late=ช็อกโกเลต	1	0.14				
co	ซ	**co**elacanth=ซีลาแคนท์	2	0.27				
cu	ก	e**cu**ador=เอกวาดอร์	5	0.68				
cr	ก	con**cr**ete=คอนกรีต	5	0.68				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
cl	ก	wy**cl**iffe= วิกลิฟฟ์	6	0.82				
cไม่สอดคล้อง			24	3.26			1	0.58
		C+(e,I,y) (244)				C (47)		
ce (136)	ซ	**ce**ll= เซลล์	123	50.41	ซ	Gree**ce**= กรีซ	43	91.49
ci (75)	ซ	**ci**gar=ซิการ์	64	26.23				
ci (75)	ช	gla**ci**er= เกลเชียร์	7	2.87				
cy (37)	ซ	**cy**clone= ไซโคลน	37	15.16				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
cสอดคล้อง			231	94.67			43	91.49
ce	ก	soc**cer**=ซอกเกอร์	1	0.41	ซ	commer**ce**=คอมเมิร์ซ	1	2.13
ce	ค	**ce**ltic=เคลติก	3	1.23	ส	veni**ce**=เวนิส	3	6.38
ce	ซ	sili**ce**ous=ซีลิเชียส	4	1.64				
ce	#	wor**ce**ster=วุสเตอร์	1	0.41				
ci	ก	as**ci**i=แอสกี	1	0.41				
ci	ค	gas**ci**ogne=แกสคอยน์	3	1.23				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
cไม่สอดคล้อง			13	5.33			4	8.51
		CH (174)				CH (32)		
ch	ช	**Ch**icago= ชิคาโก	94	54.02	ช	Bea**ch** =บีช	24	75.00
ch	ค	**Ch**ios= คิออส	78	44.83	ก	Angio**ch**= แอนติออก	8	25.00
chสอดคล้อง			172	98.85			32	100.00
ch	กก	cze**ch**oslovakia= เช็กโกสโลวะเกีย	1	0.57				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ch	จ	man**ch**uria=แมนจูเรีย	1	0.57				
chไม่สอดคล้อง			2	1.15			0	0.00
		D (717)				D (277)		
d	ด	**d**extrin=เดกซ์ทริน	717	100.00	ด	Dea**d** Sea=เดดซี	274	98.92
dสอดคล้อง			717	100.00			274	98.92
					ต	gottwal**d**=กอตต์วาลต์	1	0.36
					#	floy**d**=ฟลอย	2	0.72
dไม่สอดคล้อง			0	0.00			3	1.08

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		F (234)				F (21)		
f	ฟ	Fox= ฟอกซ์	234	100.00	ฟ	Clifion=คลิฟตัน	20	95.24
fสอดคล้อง			234	100.00			20	95.24
					#	cardiff= คาร์ดิฟ	1	4.76
fไม่สอดคล้อง			0	0.00			1	4.76
		G+(e,i,y) (204)				G (29)		
ge (136)	จ	gestagen= เจสตาเจน	115	56.37	จ	rouge=รูจ	26	89.66

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
ge (136)	ก	forget-me-not= ฟอร์เกตมีนอต	16	7.84				
gi (59)	จ	engineer= เอนจิเนียร์	33	16.18				
gi (59)	ก	gift =กิฟต์	25	12.25				
gy (12)	จ	gyro= ไจโร	6	2.94				
gสอดคล้อง			195	95.59			26	89.66
ge	ย	georges= ยอร์ช	3	1.47	ช	georges= ยอร์ช	3	10.34
gi	ต	angioch= แอนติออก	1	0.49				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

gy	ย	e**gy**pt= อียิปต์	5	2.45				
gไม่สอดคล้อง			9	4.41			3	10.34
		G + (a,o,u,l,r) (402)				G (31)		
ga (127)	ก	**ga**laxy= กาแล็กซี	127	31.59	ก	ma**g**nesium =แมกนีเซียม	31	100.00
go (93)	ก	**go**lf =กอล์ฟ	93	23.13				
gu (43)	ก	**gu**lf = กัลฟ์	43	10.70				
gl (32)	ก	**Gl**adstone= แกลดสโตน	31	7.71				
gr (107)	ก	**gr**and=แกรนด์	107	26.62				
gสอดคล้อง		409	401	99.75			31	100.00
gl	จ	mid**gl**ey=มิจลีย์	1	0.25				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
gไม่สอดคล้อง		409	1	0.25			0	0.00
		GH (7)				GH (16)		
gh	ก	**gh**etto= เกตโต	6	85.71	ฟ	Gou**gh**=กอฟ	3	18.75
					ก	Pittsbur**gh**=พิตส์เบิร์ก	2	12.50
					#	Hu**gh**=ฮิว	11	68.75
ghสอดคล้อง			6	85.71			16	100.00
gh	ค	**gh**azal=คาซัล	1	14.29				
ghไม่สอดคล้อง		10	1	14.29			0	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		GN (4)				GN (1)		
gn	น	**gn**eiss=ไนส์	4	100.00	น	desi**gn**=ดีไซน์	1	100.00
gnสอดคล้อง			4	100.00			1	100.00
gnไม่สอดคล้อง			0	0.00			0	0.00
		H (388)				H (7)		
h	ฮ	**H**aematite=อีมาไทต์	375	96.65	ห	Jo**h**n=จอห์น	7	100.00
h	#	**h**onour= ออเนอร์	13	3.35				
hสอดคล้อง			388	100.00			7	100.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
hไม่สอดคล้อง							0	0.00
		J (91)						
j	จ	Jim=จิม	88	96.70				
jสอดคล้อง			88	96.70				
j	ย	sarajevo=ซาราเยโว	3	3.30				
jไม่สอดคล้อง			3	3.30				
		K (304)				K (42)		
k	ค	Kansas=แคนซัส	141	46.38	ก	York=ยอร์ก	42	100.00
k	ก	bunker=บังเกอร์	163	53.62				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
kสอดคล้อง			304	100.00			42	100.00
kไม่สอดคล้อง			0	0.00			0	0.00
		KH (4)				KH (1)		
kh	ค	**kh**artoum=คาร์ทูม	4	100.00	ก	Si**kh**=ซิก	1	100.00
khสอดคล้อง			4	100.00			1	100.00
khไม่สอดคล้อง			0	0.00			0	0.00
		L (1489)				L (239)		
l	ล	**l**ocket=ล็อกเกต	1475	99.06	ล	She**ll**=เชลล์	229	95.82
lสอดคล้อง			1475	99.06			229	95.82

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
l	#	villanelle=วิลลาแนล	13	0.87	#	nashville=แนชวิล	9	3.77
l	ร	catabolism=แคแทบอรึซึม	1	0.07	เ.ล	toefl=โทเฟิล	1	0.42
lไม่สอดคล้อง		1516	14	0.94			10	4.18
		M (1015)				M ()		
m	ม	**m**icro=ไมโคร	1005	99.01	ม	To**m**=ทอม	250	99.21
mสอดคล้อง			1005	99.01			250	99.21
m	#	microgram**m**e=ไมโครกรัม	10	0.99	ต	communis**m**=คอมมิวนิสต์	2	0.79

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
mไม่สอดคล้อง			10	0.99			2	0.79
		N (1053)				N (697)		
n	น	nucleus= นิวเคลียส	1053	100.00	น	cyclone= ไซโคลน	697	100.00
n สอดคล้อง			1053	100.00			697	100.00
n ไม่สอดคล้อง			0	0.00			0	0.00
		P (535)				P (32)		
p	พ	paraola= พาราโบลา	305	57.01	ป	capsule= แคปซูล	31	96.87
p	ป	superman= ซูเปอร์แมน	230	42.99				

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
pสอดคล้อง			535	100.00			31	96.87
					บ	sloop=สลุบ	1	3.13
pไม่สอดคล้อง		547	0	0.00			1	3.13
		PH (131)				PH (16)		
ph	ฟ	phosphorous=ฟอสฟอรัส	131	100.00	ฟ	graph=กราฟ	16	100.00
phสอดคล้อง			131	100.00			16	100.00
phไม่สอดคล้อง		131	0	0.00			0	0.00
		Q (46)				Q (1)		
q	ก	Qatar=กาตาร์	7	15.22	ก	Iraq=อิรัก	1	100.00

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
q	ค	Liquor=ลิเคอร์	39	84.78				
qสอดคล้อง			46	100.00			1	100.00
qไม่สอดคล้อง			0	0.00			0	0.00
		R (1695)				R (421)		
r	ร	radium=เรเดียม	1686	99.47	ร	barley=บาร์เลย์	416	98.81
rสอดคล้อง			1686	99.47			416	98.81
r	ล	porarise=โพลาไรส์	2	0.12	#	torr=ทอร์	5	1.19

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
r	#	magnetostriction=แมกนีโทสติกชัน	7	0.41				
rไม่สอดคล้อง			9	0.53			5	1.19
		RH (22)				RH (1)		
rh	ร	**rh**odonite=โรโดไนต์	22	100.00	ห	myr**rh**=เมอรห์	1	100.00
rhสอดคล้อง			22	100.00			1	100.00
rhไม่สอดคล้อง			0	0.00			0	0.00
		S (2336)				S (425)		
s	ซ	**s**ilicon=ซิลิคอน	1261	53.98	ส	Lago**s**=ลากอส	395	92.94

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
s	ส	Sweden=สวีเดน	1037	44.39				
s	ช	Asia=เอเชีย	30	1.29				
sสอดคล้อง			2328	99.66			395	92.94
s	#	gross=กรอส	8	0.34	#	nabbes=แนบบ์	29	6.82
					ล	mollisols=มอลลิซอลล์	1	0.24
sไม่สอดคล้อง			8	0.34			30	7.06
		SC (2)				SC (1)		
sc	ซ	scene=ซีน	1	50.00	สก	disc=ดิสก์	1	100.00
scสอดคล้อง			1	50.00			1	100.00

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
sc	ซ	con**sc**ience =คอนเชียนซ์	1	50.00				
scไม่สอดคล้อง			1	50.00			0	0.00
		SCH (10)						
sch	ซ	**sch**ism=ซิซึม	9	90.00				
sch	ซ	**sch**eelite= ซีไลต์	1	10.00				
sch สอดคล้อง			10	100.00				
schไม่สอดคล้อง			0	0.00				
		SH (192)				SH (17)		

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
sh	ช	**sh**amal=ชามาล	190	98.96	ช	har**sh**=ฮาร์ช	17	100.00
shสอดคล้อง			190	98.96			17	100.00
sh	จ	**sh**im=จีม	1	0.52				
sh	ส	bi**sh**kek=บิสเคค	1	0.52				
shไม่สอดคล้อง			2	1.04			0	0.00
		SK				SK (4)		
sk					สก	ta**sk**=ทาสก์	4	100.00
skสอดคล้อง							4	100.00
skไม่สอดคล้อง							0	0.00

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		SM				SM		
sm					ซึม	protopla**sm** = โพรโทพลาซึม	18	85.71
smสอดคล้อง							18	85.71
					สต	commun**ism** = คอมมิวนิสต์	2	9.52
					สม	popul**ism**= ปอปปูลิสม์	1	4.76
smไม่สอดคล้อง							3	14.29

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		T (1577)				T (307)		
t	ท	Tasmania=แทสเมเนีย	476	30.18	ต	Kuwait=คูเวต	296	96.42
t	ต	intercom=อินเตอร์คอม	1043	66.14				
t	ช	nation=เนชัน	49	3.11				
tสอดคล้อง			1568	99.43			296	96.42
t	จ	stuart=สจวต	9	0.57	#	spodument=สปอดูมีน	4	1.30
					ซ	covalent=โคเวเลนซ์	1	0.33
					ท	tent=เต็นท์	6	1.95
tไม่สอดคล้อง			9	0.57		318	11	3.58

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
		TH (126)				TH (45)		
th	ท	**th**orium=ทอเรียม	124	98.41	ท	zeni**th**=เซนิท	44	97.78
thสอดคล้อง			124	98.41			44	97.78
th	ธ	eleu**th**era=อีลิวเธอรา	2	1.59	ต	kenne**th**=เคนเนต	1	2.22
thไม่สอดคล้อง			2	1.59			1	2.22
		THM				THM (4)		
thm					ทึม	biorhy**thm**=ไบโอริทึม	4	100.00
thm สอดคล้อง							4	100.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
thmไม่สอดคล้อง							0	0.00
		V (277)				V (15)		
v	ว	volt=โวลต์	277	100.00	ฟ	love=เลิฟ	12	80.00
					ฟว	Livingstone =ลิฟวิงสโตน	3	20.00
vสอดคล้อง			277	100.00			15	100.00
vไม่สอดคล้อง			0	0.00			0	0.00
		W (261)				W (34)		
w	ว		256	98.08	ว	cowboy= คาวบอย	28	82.35
wสอดคล้อง			256	98.08			28	82.35

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
w	#	mosco**w**=มอสโก	4	1.53	ร์	facebo**w**=เฟซโบร์	1	2.94
w	ย	le**w**iston=ลุยสตัน	1	0.38				
wไม่สอดคล้อง		261	5	1.92			1	2.94
		WH (12)						
wh	ว	**Wh**ite=ไวต์	11	91.67				
wh	ฮ	**Wh**ewell=ฮิวเอลล์	1	8.33				
whสอดคล้อง			12	100.00				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
whไม่สอดคล้อง			0	0.00				
		X (7)				X (38)		
x	ซ	**X**enon=ซีนอน	7	100.00	กซ	ony**x**=โอนิกซ์	37	97.37
xสอดคล้อง			7	100.00			37	97.37
					ก	orthodo**x**=ออร์ทอดอก	1	2.63
xไม่สอดคล้อง			0	0.00			1	2.63
		Y (71)				Y (111)		
y	ย	**Y**ale=เยล	71	100.00	ย	ke**y**=คีย์	111	100.00
yสอดคล้อง			71	100.00			111	100.00

พยัญชนะภาษาอังกฤษ	พยัญชนะต้น				พยัญชนะท้าย			
	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ	ถอดอักษรเป็นพยัญชนะภาษาไทย	ตัวอย่าง	ถอดอักษรได้จำนวน	คิดเป็นร้อยละ
yไม่สอดคล้อง			0	0.00			0	0.00
		Z (133)				Z (24)		
z	ซ	zone=โซน	133	100.00	ซ	Vaduz=วาดุซ	22	91.67
zสอดคล้อง			133	100.00			22	91.67
					#	agassiz=แอกาซี	2	8.33
zไม่สอดคล้อง			0	0.00			2	8.33

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาววัลย์วรา ไชยฤกษ์ เกิดวันที่ 15 พฤษภาคม พ.ศ. 2523 สำเร็จการศึกษาระดับปริญญาตรีอักษรศาสตรบัณฑิต สาขาภาษาเกาหลี ภาควิชาภาษาตะวันตก จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2543 และเข้าศึกษาต่อในหลักสูตรอักษรศาสตรมหาบัณฑิต ภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ. 2544

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย