

การออกแบบขั้นตอนวิธีตัวแปลร้อยกรองไทย เป็นภาษาอังกฤษแบบคงฉันทลักษณ์ Designing of an Algorithm for Thai to English Poem Translator with Prosody Keeping

สัจจาภรณ์ ไวจรรยา (Sajjaporn Waijanya)* และ อนิราช มิ่งขวัญ (Anirach Mingkhwan)**

บทคัดย่อ

การสร้างเครื่องแปลร้อยกรองจากภาษาต้นฉบับไปเป็นอีกภาษาหนึ่งมีความแตกต่างจากการสร้างเครื่องแปลภาษาทั่วๆ ไป เนื่องจากร้อยกรองเป็นงานเขียนที่มีฉันทลักษณ์ ร้อยกรองไทยกำหนดจำนวนพยางค์ สัมผัสระหว่างบท สัมผัสระหว่างวรรค ซึ่งเนื้อความในวรรคอาจไม่ครบถ้วนตามวากยสัมพันธ์ การใช้เครื่องมือแปลภาษาทั่วไปเพื่อแปลร้อยกรองไทยไปเป็นภาษาอังกฤษไม่สามารถแปลงาน ร้อยกรองให้คงฉันทลักษณ์และความหมายเดิมได้ ต้องปรับแต่งเครื่องมือแปลภาษา งานวิจัยนี้จึงนำเสนอการออกแบบขั้นตอนวิธีตัวแปลร้อยกรองไทยเป็นภาษาอังกฤษแบบคงฉันทลักษณ์ โดยใช้ร้อยกรองประเภทกลอนสุภาพ ดอกสร้อย และ สักวา และออกแบบการวัดผลตัวแปลร้อยกรองโดยวัดเปอร์เซ็นต์ความถูกต้อง เพื่อให้เกิดการต่อยอดการสร้างเครื่องมือแปลร้อยกรองไทยที่หลากหลาย ทั้งยังส่งเสริมการเผยแพร่วรรณศิลป์ไทยให้เป็นที่สนใจในระดับสากลมากยิ่งขึ้น

คำสำคัญ: การแปลร้อยกรองไทย เครื่องมือแปลภาษา ตัวแปลร้อยกรอง

Abstract

The poem translation from original language to another is very different from general machine translator because the poem is written with prosody. Thai poem is composed with set of syllables which to be rhyme with stanza and line and the text in paragraph of poem may not be complete syntax.

* ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

** ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Thus, general machine translator absolutely cannot translate Thai poem to English poem which prosody keeping and meaning are still the same. This research presents designing of an algorithm for Thai poem to English poem translator with prosody keeping to translate 3 poem types are Klon Supaap, Doksoi and Sakkawa to English. Measurement designs in this research are the accuracy percentage defined with prosody. Therefore, this research will be the new innovation of the machine translator for poem, and encourage Thai art created language to the global as well.

Keyword: Thai Poem translation, Machine Translator, Poem translator.

1. บทนำ

งานเขียนร้อยกรองประเภทกลอนโดยเฉพาะกลอนสุภาพ นอกจากจะถ่ายทอดความนึกคิดของผู้เขียนสู่ผู้อ่านแล้ว ยังสร้างความประทับใจในความงดงาม ไพเราะ และลึกซึ้งในถ้อยคำ การประพันธ์ร้อยกรองยังเป็นสุนทรียศาสตร์และศิลปะซึ่งเป็นมรดกทางวัฒนธรรมที่ต้องรักษาไว้ ร้อยกรองของหลายชนชาติที่ได้รับการแปลจากภาษาต้นฉบับเป็นภาษาอังกฤษ นับเป็นทางหนึ่งในการเผยแพร่วัฒนธรรมให้เข้าถึงนานาชาติอารยประเทศ ร้อยกรองของไทยคือวรรณศิลป์ที่ถ่ายทอดความเป็นไทยเช่นกัน พระวรราชเธอ พระองค์เจ้าเปรมบุรฉัตร ทรงแปลร้อยกรองเรื่อง “พระอภัยมณี” ที่เป็นบทประพันธ์ของ “พระสุนทรโวหาร (สุนทรภู่)” กวีเอกของ

ไทย ผู้ได้รับการยกย่องให้เป็นกวีเอกของโลก โดยทรงแปลเป็นภาษาอังกฤษทั้งเรื่อง เมื่อปี พ.ศ. 2495 [1] เพื่อเผยแพร่ให้นานาประเทศได้ชื่นชม และศึกษาการแปลร้อยกรองจากต้นฉบับเป็นภาษาอังกฤษ หรือจากภาษาอังกฤษกลับมาเป็นต้นฉบับนั้น เป็นแขนงหนึ่งของการแปลภาษาด้วยคอมพิวเตอร์ ที่นักวิจัยผู้สร้างตัวแปลภาษาให้ความสนใจเนื่องจากร้อยกรอง มีจังหวะ มีรูปแบบในการร้อยเรียงคำที่เรียกว่า “ฉันทลักษณ์”

งานวิจัยต่างประเทศจำนวนหนึ่งได้สร้างตัวแปลภาษาจากร้อยกรองเช่น จีน ญี่ปุ่น อินเดีย สเปน อิตาลี ไปเป็นภาษาอังกฤษ และแปลจากภาษาอังกฤษ กลับไปเป็นภาษาเหล่านั้น ทั้งยังพบว่ามีงานวิจัยอีกกลุ่มที่สร้างบทกลอนญี่ปุ่น (ไฮกุ) สร้างกลอนจีน ในรูปของภาษาอังกฤษ ซึ่งทำให้กลอนของแต่ละประเทศเหล่านั้นถูกนำเสนอในระดับสากล เป็นการส่งเสริมและเผยแพร่วัฒนธรรมโดยผ่านเทคโนโลยี แต่กระนั้นกลับพบว่าไม่มีงานวิจัยใดที่สร้างตัวแปลร้อยกรองไทยไปเป็นภาษาอังกฤษ เจกเช่นร้อยกรองภาษาอื่นๆ ดังนั้นงานวิจัยนี้จึงนำเสนอแนวคิด การออกแบบ และขั้นตอนวิธีตัวแปลร้อยกรองไทยเป็นภาษาอังกฤษแบบคงฉันทลักษณ์ ร้อยกรองไทยและคงความหมาย โดยร้อยกรองที่ใช้ในงานวิจัยนี้คือร้อยกรองประเภท “กลอนสุภาพ” “กลอนดอกสร้อย” และ “กลอนสักรา” ที่มีความไพเราะ และความเรียบง่ายต่อการสื่อความ ทั้งยังส่งเสริมให้รูปแบบและความสละสลวยของกลอนไทยได้รับการเผยแพร่มากขึ้นในระดับสากล ซึ่งนับเป็นจุดเริ่มต้นของร้อยกรองไทยในงานวิจัยด้านการสร้างตัวแปลภาษา

2. งานวิจัยทางคอมพิวเตอร์กับบทร้อยกรอง

งานวิจัยทางเทคโนโลยีคอมพิวเตอร์ที่เกี่ยวข้องกับบทร้อยกรอง ไม่ว่าจะเป็นกระบวนการในการสร้างร้อยกรองด้วยคอมพิวเตอร์ หรือกระบวนการในการแปลร้อยกรอง นับเป็นงานที่อยู่ในกลุ่มเทคโนโลยีการประมวลผลภาษาธรรมชาติ (Natural language processing) และใช้เทคโนโลยีที่เรียกว่าเครื่องแปลภาษา (Machine Translation) เป็นหัวใจหลักในการประมวลผล ซึ่งในการศึกษางานวิจัยที่เกี่ยวข้องนี้ สามารถแบ่งกลุ่มของงานวิจัยที่ได้ศึกษาออกเป็น 3 กลุ่ม คือ 1) งานวิจัยที่เกี่ยวข้องกับเรื่องการตรวจสอบฉันทลักษณ์ของร้อยกรอง 2) งานวิจัยที่เกี่ยวข้องกับการสร้างบทร้อย

กรองด้วยคอมพิวเตอร์ 3) งานวิจัยเกี่ยวกับการแปลร้อยกรองจากภาษาหนึ่งไปอีกภาษาหนึ่ง ด้วยคอมพิวเตอร์

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบฉันทลักษณ์คำประพันธ์ ออกแบบภาษาเชิงไวยากรณ์สำหรับคำประพันธ์ โดยการสร้างกฎสำหรับคำประพันธ์แต่ละชนิดและจัดทำเครื่องมือเพื่อช่วยในการแต่งคำประพันธ์ ที่นำเสนอการแต่งคำประพันธ์ในรูปแบบที่น่าสนใจและง่ายขึ้น ซึ่งมีทั้งนักวิจัยไทย [2, 3] และต่างชาติ [4] ผลงานวิจัยในแขนงนี้

งานวิจัยที่เกี่ยวข้องกับการสร้างบทร้อยกรองด้วยคอมพิวเตอร์ ผู้วิจัยใช้แนวทางของปัญญาประดิษฐ์ และเทคโนโลยีของเว็บ 2.0 ในการสร้างกลอนไฮกุแบบอัตโนมัติ โดยใช้เนื้อหาจากบล็อก [5] และอีกงานวิจัยหนึ่งนำเสนอเทคนิค A*Search หนึ่งในเทคนิคของเครื่องแปลภาษา (MT) ควบคู่กับ template ในการสร้างบทร้อยกรองเทียบกับ เทคนิคคลังข้อมูล [6]

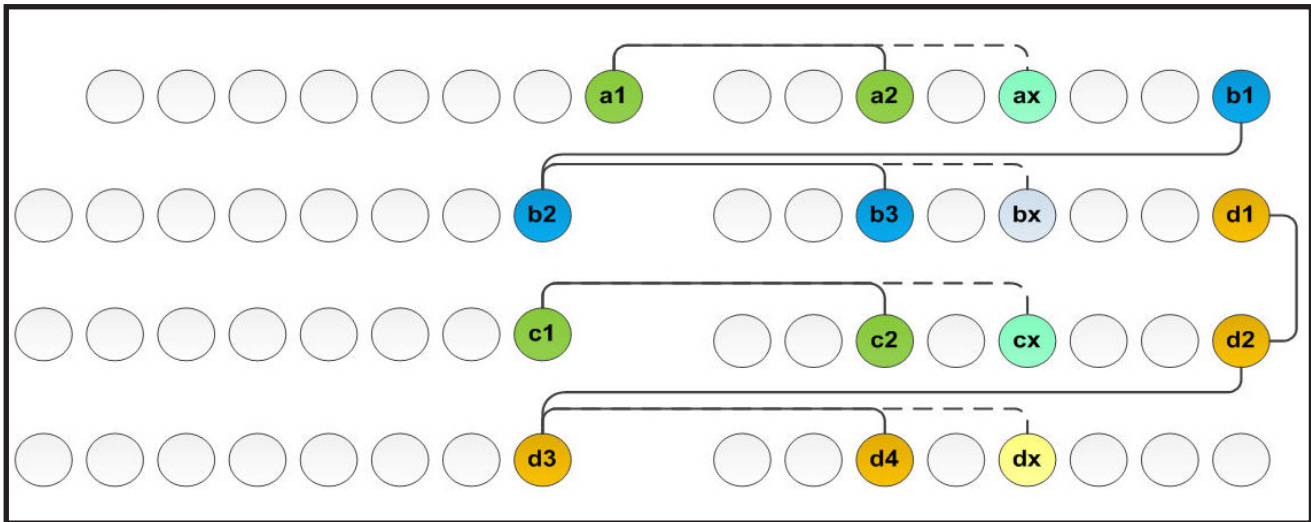
งานวิจัยเกี่ยวกับการแปลร้อยกรองจากภาษาหนึ่งไปอีกภาษาหนึ่ง ด้วยคอมพิวเตอร์ ผู้วิจัยใช้ BLEU (Bilingual Evaluation Understudy) ซึ่งเป็นกระบวนการที่นิยมมากในการสร้างตัวแปลภาษา ซึ่งหาค่า N-gram ระหว่าง Test-text และ Reference text ในการแปลบทกวีจีนที่ชื่อว่า Yellow Crane Tower ของราชวงศ์ถัง เป็นภาษาอังกฤษ [7] และยังพบว่า Team MT ของบริษัทกูเกิ้ลได้ให้ความสนใจในการแปลบทร้อยกรองด้วยการนำเสนอขั้นตอนวิธีใหม่ โดยใช้ Stress Pattern ร่วมกับการใช้วิธีการอื่นๆ ในกระบวนการแปลภาษา [8] งานวิจัยที่เกี่ยวข้องกับร้อยกรองไทยมีเพียงงานวิจัยในกลุ่มการตรวจสอบฉันทลักษณ์เท่านั้น ส่วนในอีก 2 กลุ่ม ไม่พบว่ามีการพัฒนาสำหรับร้อยกรองไทยแต่อย่างใด

3. ฉันทลักษณ์และข้อกำหนดของกลอนในงานวิจัย

3.1 ฉันทลักษณ์และข้อกำหนดทั่วไปของกลอนสุภาพ

กลอนสุภาพหรือกลอนแปด 1 บท ประกอบด้วย 4 วรรค คือ วรรคที่ 1 เรียกว่าวรรคสลับ วรรคที่ 2 เรียกว่าวรรครับ วรรคที่ 3 เรียกว่าวรรครอง และวรรคที่ 4 เรียกว่าวรรคส่ง โดยมีลักษณะบังคับที่สำคัญ 3 ประการ [9] คือ

- 1) บังคับจำนวนพยางค์ ในแต่ละวรรคใช้ได้ 8 พยางค์ ทั้งนี้ อนุญาตให้ใช้ 7 หรือ 9 พยางค์ได้
- 2) บังคับเสียงท้ายวรรค คือ วรรคที่ 1 ใช้ได้ทุกเสียง แต่



ภาพที่ 1 แผนผังกลอนสุภาพ 2 บท

ไม่นิยมเสียงสามัญ วรรคที่ 2 ห้ามใช้เสียง ตรี และ สามัญ วรรคที่ 3 และ 4 ห้ามใช้เสียง เอก โท และ จัตวา

3) บังคับสัมผัส มีลักษณะดังภาพที่ 1 คือ สัมผัสนอก มี 2 ตอน ตอนที่ 1 สัมผัสของพยางค์ท้ายวรรคที่ 1 กับพยางค์ที่ 3 หรือพยางค์ที่ 5 ของวรรคที่ 2 ตอนที่ 2 คือสัมผัสของพยางค์ท้ายวรรคที่ 2 กับพยางค์ท้ายวรรคที่ 3 และ พยางค์ที่ 3 หรือพยางค์ที่ 5 ของวรรคที่ 4

สัมผัสเชื่อมบท หมายถึง พยางค์ท้ายวรรคที่ 4 ที่ต้องส่งสัมผัสไปยังพยางค์ท้ายวรรคที่ 2 ของบทถัดไป นอกจากข้อกำหนดของสัมผัสข้างต้น กลอนที่มีคุณภาพและมีความไพเราะ ไม่ควรเขียนในลักษณะซึ่งสัมผัส หรือใช้สัมผัสซ้ำ

การชิงสัมผัส หมายถึง ในวรรคมีการใช้พยางค์ที่มีเสียงสระเสียงเดียวกันกับพยางค์ในตำแหน่งที่รับสัมผัส เช่น จากภาพที่ 1 a1 ส่งสัมผัสไปยัง a2 หรือ ax หากมีเสียงรับสัมผัสในตำแหน่งก่อน a2 จะเป็นการชิงตำแหน่งรับสัมผัส ส่วนกรณีที่พบพยางค์ที่มีเสียงซ้ำกับพยางค์ส่งสัมผัส แม้ไม่นับเป็นชิงสัมผัส แต่ก็ทำให้กลอนด้อยความไพเราะ

สัมผัสซ้ำ หมายถึง พยางค์ส่งและพยางค์รับสัมผัสเป็นพยางค์ที่มีพยัญชนะและสระเป็นเสียงเดียวกันเช่น วรรคที่ 1 ลงท้ายด้วยคำว่า “เพลินตา” พยางค์ a1 คือ “ตา” ถ้าวรรคที่ 2 ขึ้นด้วยคำว่า “อนัตตา” พยางค์ที่รับสัมผัส a2 คือ “ตา” เป็นสัมผัสซ้ำ

3.2 จันท์ลักษณะและข้อกำหนดทั่วไปของกลอนสักรา

กลอนสักรา 1 บทมี 4 คำกลอน (8 วรรค) จังหวะแรกของวรรคขึ้นต้นบทขึ้นต้นด้วยคำว่า “สักรา” คำสุดท้ายของบท

ลงท้ายด้วย “เอ๋ย” สัมผัสบังคับเป็นเหมือนกับกลอนสุภาพ

3.3 จันท์ลักษณะและข้อกำหนดทั่วไปของกลอนดอกสร้อย

กลอนดอกสร้อย 1 บทมี 4 คำกลอน (8 วรรค) วรรคแรกขึ้นต้นใช้ 4 คำ โดยเอา 2 คำมาตั้งแล้วซ้ำคำหน้าโดยแทรกคำว่า “เอ๋ย” ระหว่างกลาง เช่น “น้ำเอ๋ยน้ำคำ” คำสุดท้ายของบทใช้คำว่า “เอ๋ย” สัมผัสบังคับเหมือนกับกลอนสุภาพ

3.4 จันท์ลักษณะและข้อกำหนดที่ใช้ในงานวิจัย

ในงานวิจัยนี้ใช้เฉพาะจันท์ลักษณะ และข้อกำหนดบางข้อ โดยเลือกข้อที่สำคัญ คือ

- 1) บังคับจำนวนพยางค์
- 2) บังคับสัมผัส วัตถุประสงค์ สัมผัสนอก สัมผัสเชื่อมบท และสัมผัสซ้ำ

นอกจากนี้ เพื่อให้คงจันท์ลักษณะเมื่อกลอนที่นำมาแปลเป็นกลอนดอกสร้อย หรือสักรา จึงกำหนดให้ “สักรา” ทับศัพท์เป็น “Sakkawa” “เอ๋ย” ทับศัพท์เป็น “Oy” และ “เอ๋ย” ทับศัพท์เป็น “Oei”

3.5 จันท์ลักษณะและข้อกำหนดที่ละเว้น

เนื่องจากงานวิจัยนี้ออกแบบให้กลอนที่นำมาทดสอบเป็นกลอนไทยที่ถูกต้องตามจันท์ลักษณะ เมื่อผ่านกระบวนการแปลภาษาเป็นภาษาอังกฤษแล้วจึงตรวจสอบจันท์ลักษณะและละเว้นจันท์ลักษณะบางข้อ ดังนี้

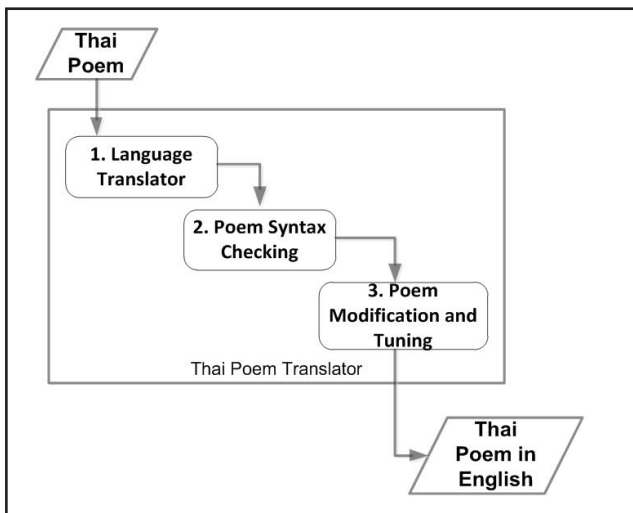
- 1) ละเว้นบังคับเสียงท้ายวรรค
- 2) ละเว้นบังคับชิงสัมผัส

จันท์ลักษณะที่ละเว้นในงานวิจัยนี้ได้รับการรับรองให้

ละเว้นได้โดยผู้เชี่ยวชาญจากสมาคมกวีร่วมสมัย เนื่องจากบทกลอนที่นำมาทดสอบไม่ใช่บทกลอนในการประกวด และด้วยข้อจำกัดของเสียงในภาษาอังกฤษ ซึ่งผู้เชี่ยวชาญเห็นว่าข้อบังคับที่ละเว้นไม่ก่อให้เกิดความเสียหายแต่อย่างใด

4. การวิเคราะห์ และออกแบบ

เนื่องจากงานวิจัยนี้ออกแบบขั้นตอนวิธีการแปลกลอนจากภาษาไทยเป็นภาษาอังกฤษ โดยมุ่งออกแบบเพื่อสร้างเครื่องมือแปลร้อยกรองจากภาษาไทยไปเป็นภาษาอังกฤษแบบคงความหมาย และคงฉันทลักษณ์ให้ได้มากที่สุด ซึ่งมีกรอบแนวคิด และการออกแบบองค์ประกอบแสดงในรูปแบบโมเดลกระบวนการ ดังภาพที่ 2



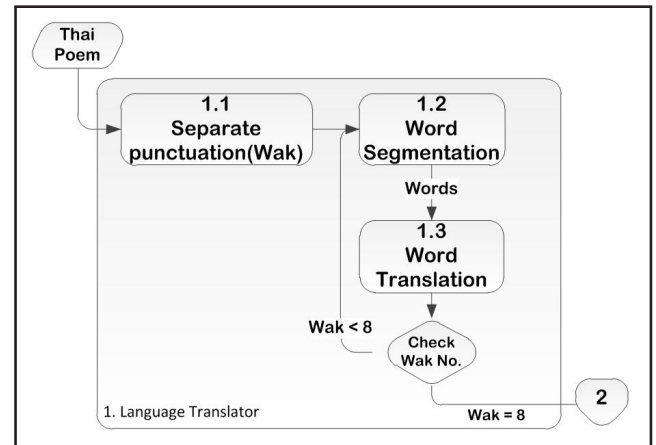
ภาพที่ 2 การออกแบบองค์ประกอบของโมเดลกระบวนการ

โดยทั้ง 3 กระบวนการ อธิบายได้ดังนี้

4.1 กระบวนการแปลความหมาย

การออกแบบขั้นตอนและกระบวนการแปลความหมายแบ่งเป็น 3 กระบวนการย่อย คือ 1.1) การแบ่งวรรค (Separate punctuation (Wak)) 1.2) การตัดคำ (Word Segmentation) และ 1.3) การแปลความหมาย (Word Translation) แสดงดังภาพที่ 3

ข้อมูลที่เป็นอินพุตในงานวิจัยนี้ คือ ร้อยกรองไทยประเภท กลอนแปด 2 บท สักวา และดอกสร้อย ซึ่งมีความยาว 8 วรรค ผ่านกระบวนการที่ 1.1) เพื่อแบ่งวรรค จากนั้นแต่ละวรรคเข้าสู่กระบวนการ 1.2) เพื่อตัดคำโดยใช้อินเทอร์เฟซของตัวตัดคำภาษาไทย และเข้าสู่กระบวนการที่ 1.3) เพื่อแปลความหมายของคำ โดยระบบจะดำเนินการ



ภาพที่ 3 การออกแบบกระบวนการแปลความหมาย

กระบวนการที่ 1) กระบวนการแปลความหมาย จนครบทั้ง 8 วรรค จากนั้นส่งข้อมูลจากกระบวนการที่ 1) ไปยังกระบวนการที่ 2) การตรวจสอบฉันทลักษณ์ ต่อไป และเพื่อให้เห็นภาพที่ชัดเจนขึ้น ผู้วิจัยจึงนำเสนอการทำงานของแต่ละกระบวนการพร้อมทั้งตัวอย่างประกอบ ดังนี้

ตัวอย่างบทร้อยกรอง สำหรับตัวแปลร้อยกรอง

โธ่ละเห่ฟุ่มพักด้วยรักเจ้า
เมื่อยามเฝ้าแม่ถนอมกล่อมจอมขวัญ
เอื้ออาทรอ่อนร้าย แม่คลายพลัน
ภัยกังกันแม่กล้าฟันฝ่าไป
โธ่ละเห่ที่ฟุ่มพักถูกรักแม่
จะแน่วแน่มั่นเป็นคนดีมีผลอไผ่
เพื่อทดแทนคุณความรักจากดวงใจ
ร้อยมัลลย์อักษรบูชาเอย

4.1.1 การแบ่งวรรค

การแบ่งวรรค จะแบ่งตามตัวจบวรรคโดยงานวิจัยนี้ อ่านรหัสแอสกีเครื่องหมาย “\t” เป็นตัวจบวรรค เช่น โธ่ละเห่ฟุ่มพักด้วยรักเจ้า \t จากนั้นบันทึกวรรคที่แบ่งแล้วลงในอาร์เรย์ตามตำแหน่ง 1 ถึง 8 และนำข้อมูลเข้าสู่กระบวนการที่ 1.2 และ 1.3 ที่ละวรรค จนครบ

4.1.2 การตัดคำ

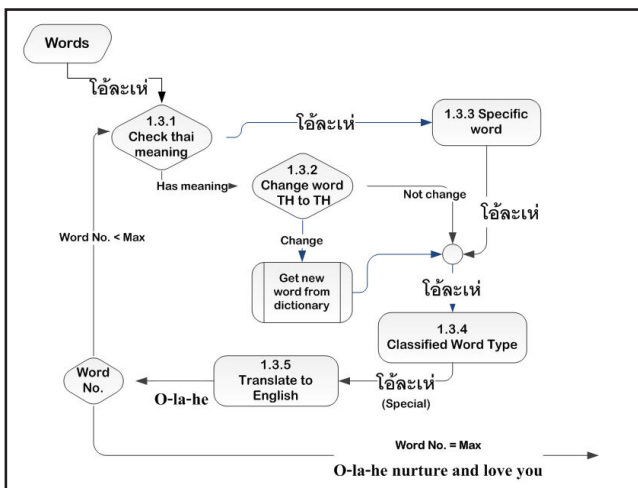
การตัดคำในงานวิจัยนี้ ใช้อินเทอร์เฟซตัวตัดคำของมหาวิทยาลัยเกษตรศาสตร์ ชื่อ KU-CUT [10] ซึ่งสามารถระบุประเภทของคำที่ตัดได้ ตัวอย่าง เช่น โธ่ละเห่ฟุ่มพักด้วยรักเจ้า ตัดคำได้เป็น โธ่ละเห่|ฟุ่มพัก|ด้วย|รัก|เจ้า แต่เนื่องจากอินเทอร์เฟซที่ใช้ในกระบวนการตัดคำสำหรับงานวิจัยนี้เป็นโปรแกรมที่ใช้ในการตัดคำทั่วไป ดังนั้นผู้วิจัยจึงออกแบบกฎ

เพื่อปรับผลลัพธ์จากการตัดคำอีกครั้ง โดยการอ้างอิงบริบททางหลักภาษาไทย

ในวรรคนี้ประเภทของคำแต่คำคือ โอ้<กรรมกริยา>|ละ<กริยา>|เห่<กริยา>|พุ่มพัก<กริยา>|ด้วย<สันธาน>|รัก<กริยา>|เจ้า<สรรพนาม> แม้บริบทประโยคภาษาไทย คือ “ประธาน-กริยา-กรรม” แต่ในภาษากลอน ประโยคอาจจะลดรูปได้ ซึ่งการจำแนกประเภทของคำจะถูกใช้ประโยชน์ในการรวมคำ เช่น คำที่เกิดจากกริยาเรียง หรือ กริยาผสม รวมถึงประโยคที่ถูกลดรูป เป็นต้น

4.1.3 การแปลความหมายของคำ

การออกแบบกระบวนการแปลความหมายของคำ ออกแบบให้ประกอบด้วยขั้นตอนย่อยๆ อีก 5 ขั้นตอน คือ 1.3.1) ตรวจสอบความหมายของคำ 1.3.2) การแปลคำไทย



ภาพที่ 4 การออกแบบกระบวนการย่อยที่

1.3) การแปลความหมายของคำ

โอ้ละเห่ เป็นคำที่ไม่มีความหมายตามพจนานุกรม แต่เป็นเสียงที่ใช้ในบทเพลงกล่อมเด็ก แทนเสียงกล่อมของแม่ม ดังนั้นคำว่า โอ้ละเห่ จะถูกระบุว่าเป็นคำเฉพาะเพื่อให้ระบบแปลแบบทัวตัพทในขั้นตอนกระบวนการที่ 1.3.5 ซึ่งกระบวนการที่ 1.3 จำลองตัวอย่างข้อมูลกับโมเดลได้ดังนี้ ตัวอย่าง วรรคที่ 1 และ วรรคที่ 2 เมื่อผ่านกระบวนการที่ 1.3

โอ้ละเห่|พุ่มพัก|ด้วย|รัก|เจ้า->โอ้ละเห่|พุ่มพักด้วยรักเจ้า
O-la-he nurture and love you
เมื่อ|ยาม|เยาว์|แม่|ถนอม|กล่อม|จอมขวัญ-> เมื่อ|ยาม
เยาว์|แม่|ถนอม|กล่อม|จอมขวัญ

When young | mother nurture care honey

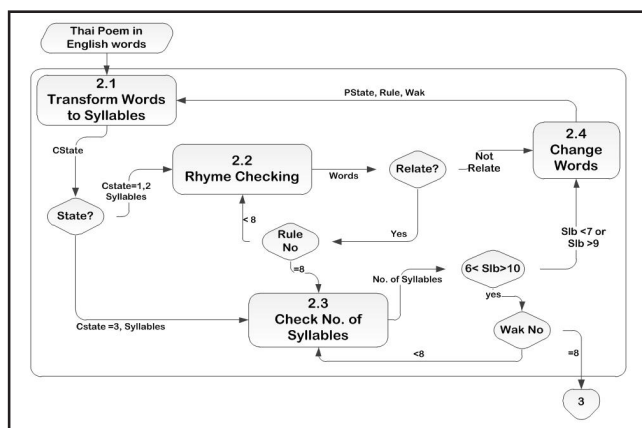
4.2 กระบวนการตรวจสอบฉันทลักษณ์

จากงานวิจัยเดิมของผู้วิจัย [3] ที่ได้สร้างภาษาไวยากรณ์ (BNF) สำหรับตรวจทานฉันทลักษณ์ของร่ายกรองประเภทกลอนแปดไว้แล้ว ในงานวิจัยนี้จึงนำไวยากรณ์เดิมมาปรับปรุงให้สอดคล้องกับการแปลร่ายกรองจากภาษาไทยให้เป็นภาษาอังกฤษ ปรับปรุงการอ่านเสียงพยางค์จากเดิมที่เป็นรูปแบบการอ่านเสียงพยางค์ไทยเป็นการอ่านเสียงพยางค์ของคำในภาษาอังกฤษ โดยเปลี่ยนการอ่านออกเสียงพยางค์ด้วยตารางสัทอักษรถ่ายไปเป็นตารางสัทอักษรภาษาอังกฤษ การตรวจสอบจะนับจำนวนพยางค์ในวรรคให้อยู่ระหว่าง 7 ถึง 9 พยางค์ และตรวจสอบสัมผัสตามภาพที่ 1 จากโครงสร้างและตำแหน่งสัมผัสในภาพที่ 1 กำหนดเป็นกฎของการตรวจสอบสัมผัสของกลอน 2 บทได้ 14 ข้อ คือ

- R1 สัมผัสระหว่าง a1 และ a2 หรือ a1 และ ax
- R2 สัมผัสระหว่าง b1 และ b2
- R3 สัมผัสระหว่าง b1 และ b3 หรือ b1 และ bx
- R4 สัมผัสระหว่าง b2 และ b3 หรือ b2 และ bx
- R5 สัมผัสระหว่าง b1, b2 และ b3 หรือ b1, b2 และ bx
- R6 สัมผัสระหว่าง c1 และ c2 หรือ c1 และ cx
- R7 สัมผัสระหว่าง d1 และ d2
- R8 สัมผัสระหว่าง d1 และ d3
- R9 สัมผัสระหว่าง d1 และ d4 หรือ d1 และ dx
- R10 สัมผัสระหว่าง d2 และ d3
- R11 สัมผัสระหว่าง d2 และ d4 หรือ d2 และ dx
- R12 สัมผัสระหว่าง d2, d3 และ d4 หรือ d2, d3 และ dx
- R13 สัมผัสระหว่าง d3 และ d4 หรือ d3 และ dx
- R14 สัมผัสระหว่าง d1, d2, d3 และ d4 หรือ d1, d2, d3 และ dx

จากข้อกำหนดในการตรวจสอบฉันทลักษณ์ที่ได้กล่าวไว้ข้างต้น ผู้วิจัยออกแบบให้ระบบมีลำดับขั้นตอนการตรวจสอบฉันทลักษณ์เป็นกระบวนการย่อย 3 กระบวนการ คือ

- 1) กระบวนการ 2.1 แบ่งพยางค์ และกำกับเสียงอ่าน
- 2) กระบวนการ 2.2 ตรวจสอบสัมผัส
- 3) กระบวนการ 2.3 ตรวจสอบจำนวนพยางค์ แสดงโมเดลของกระบวนการที่ 2 ดังภาพที่ 5
- 1) กระบวนการที่ 2.1 แปลงคำภาษาอังกฤษที่ได้จากกระบวนการที่ 1 ให้เป็นพยางค์พร้อมกำกับเสียงอ่าน



ภาพที่ 5 การออกแบบโมเดลของกระบวนการตรวจสอบ
อันตราย

ตัวอย่างเอาต์พุตที่ได้จากกระบวนการที่ 2.1 (แสดงตัวอย่างเฉพาะวรรคที่1) คือ

O-la-he nurture and love you

“O-la-he”| nur|ture| and| love| **you**

2) กระบวนการที่ 2.2 ตรวจสอบสัมพัทธ์ตามตำแหน่งที่กำหนดโดยไล่ไปตามกฎทั้ง 14 ข้อ ตามลำดับของวรรคโดยตรวจสอบเงื่อนไขตามตำแหน่งสัมพัทธ์ของแต่ละวรรค เทียบจากกฎ 14 ข้อที่สัมพันธ์กัน หากพยางค์ใดตำแหน่งที่ตรวจสอบไม่สัมพันธ์ ระบบจะส่งค่าที่ต้องตรวจนั้นไปกระบวนการที่ 2.4 เพื่อเปลี่ยนค่านั้นทั้งค่าเป็นค่าใหม่ที่มีความหมายเดิมหรือใกล้เคียงโดยใช้คำจากพจนานุกรมมาตรฐาน และพจนานุกรมที่สร้างเป็นออนโทโลยีของคำ ซึ่งกระบวนการที่ 2.4 จะอธิบายเพิ่มเติมในลำดับถัดไป

ในภาพที่ 6 แสดงตัวอย่างการตรวจสอบสัมผัสของวรรคที่ 1 และ วรรคที่ 2 ซึ่งตรวจสอบตามกฎหมาย R1 ในกระบวนการที่ 2.2 นี้ เมื่อตรวจสอบข้อมูลตัวอย่างพบว่า “ไม่ตรงสัมผัส” จึงต้องเข้าสู่กระบวนการที่ 2.4 เพื่อหาคำศัพท์ใหม่ที่สัมผัสกันตามตำแหน่งที่ระบุโดยกฎเพื่อให้ได้ตรงตามความถูกต้องของฉันทลักษณ์

ตัวอย่างผลลัพธ์จากกระบวนการที่ 2.2 คือ

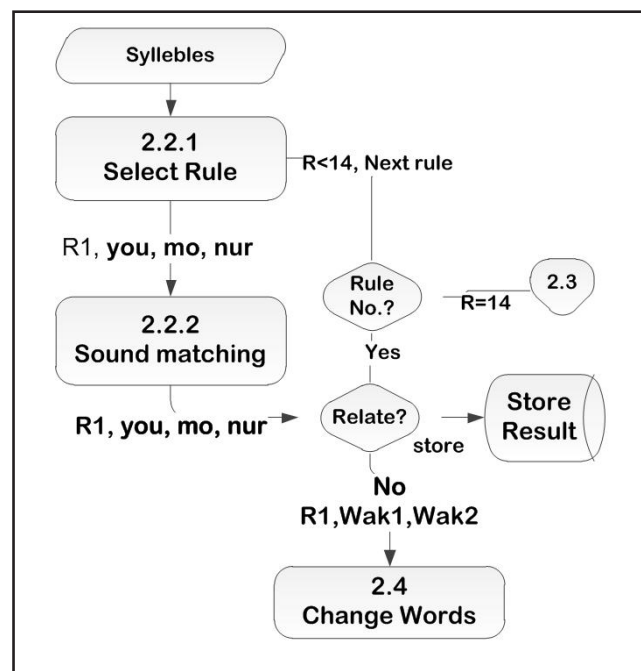
“O-la-he” to| nur|ture| and| love| **you**

“O-la-he” to nurture and love you

when| young| mo|ther nur|ture| cares ho|ney

When young mother to nurture cares honey

3) กระบวนการที่ 2.3 การตรวจสอบจำนวนพยางค์ของแต่ละวรรค คือ จำนวนพยางค์ในแต่ละวรรค ต้องมีจำนวน



ภาพที่ 6 การออกแบบโมเดลและตัวอย่างข้อมูล
การตรวจสอบสัมพัทธ์

ระหว่าง 7 ถึง 9 พยางค์ โดยหากไม่ตรงตามเงื่อนไขของ
ฉันทลักษณ์ระบบจะส่งค่าไปยังกระบวนการที่ 2.4 เพื่อ
เปลี่ยนค่า และตรวจสอบเงื่อนไขการนับจำนวนพยางค์ใหม่
โดยยกเว้นค่าที่อยู่ในตำแหน่งที่มีพยางค์รับ หรือ ส่งสัมผัส
ตัวอย่างผลลัพธ์จากกระบวนการที่ 2.3 คือ

“O-la-he” to nurture and love you

“O-la-he” to nurture and love you

When young mother **to** nurture cares **honey**

When young mother to nurture cares honey

4) กระบวนการที่ 2.4 การเปลี่ยนคำ จะเกิดขึ้นเมื่อ กระบวนการที่ 2.2 และ 2.3 รายงานความผิดพลาดในส่วน ของสัมผัส และ ส่วนของจำนวนพยางค์ โดยการเปลี่ยนคำ จะมีเงื่อนไขและขั้นตอน คือ

จากกระบวนการที่ 2.2 เมื่อพบว่า มีสัมผัสน้ำ หรือ ไม่
สัมผัสตามตำแหน่งที่ระบุในกฎที่กล่าวไว้ข้างต้น ระบบจะ
เปลี่ยนคำนั้นเป็นคำใหม่ที่มีความหมายเดิม หรือใกล้เคียง
โดยใช้คำจากพจนานุกรมมาตรฐาน พจนานุกรมศัพท์เฉพาะ
และพจนานุกรมที่สร้างเป็น ออนโทโลยีของคำ

จากกระบวนการที่ 2.3 เมื่อพบว่าจำนวนพยางค์ในวรรค
นั้นๆ มีความยาวน้อยกว่า 7 หรือ มากกว่า 9 พยางค์ ระบบ
จะดำเนินการโดยตรวจสอบ และเปลี่ยนค่าด้วย 3 เงื่อนไข

คือ

1) เปลี่ยนคำที่ละคำ โดยเปลี่ยนคำให้มีความหมายเดิม แต่มีความยาวของพยางค์ สั้นลง หรือ ยาวขึ้น

2) เลือกตัดคำที่อยู่ใกล้กันและมีความหมายในกลุ่มเดียวกันออก

3) เติมคำนำหน้า เช่น a, an, the, to เป็นต้น

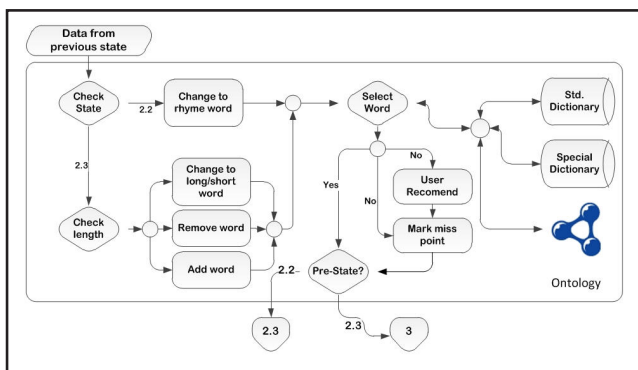
หากไม่สามารถหาคำมาทดแทนได้ระบบจะออกแบบให้

1) ผู้ใช้สามารถแนะนำคำศัพท์ได้ จากนั้น ระบบจะต้องบันทึกคำศัพท์นั้นเข้าพจนานุกรมในกรณีเป็นศัพท์ใหม่

2) ผู้ใช้สามารถเขียนบทกลอนในวรรณคดีใหม่ได้ ระบบจะต้องบันทึกคำศัพท์นั้นเข้าพจนานุกรมในกรณีพบศัพท์ใหม่

3) ผู้ใช้สามารถยืนยันการใช้คำเดิมที่ไม่สัมผัสได้

โดยระบบจะต้องบันทึกแต่ที่ระบบดำเนินการผิดพลาดไว้เพื่อสรุปผลวิจัย ซึ่งการทำงานของกระบวนการเปลี่ยนคำ แสดงดังภาพที่ 7



ภาพที่ 7 โมเดลกระบวนการเปลี่ยนคำ

ได้ผลลัพธ์ในกระบวนการที่ 2 ดังนี้

“O-la-he” to nurture and love you

When young mother to cares honey

Serious trouble mum release promptly

Mother plucky to hamper serious

“O-la-he” to nurture the child love mum my

The firmly is not unconscious

To reward your love to precious

Poem garlands letter worship “Oei”

4.3 กระบวนการปรับแต่ง

เนื่องจากงานวิจัยนี้เป็นงานวิจัยแรกที่ทำให้ความสนใจในการสร้างตัวแปลร่ายกรองไทย ซึ่งโดยทั่วไปแล้วในการแปล “คำศัพท์” และ “ความสัมพันธ์ของคำศัพท์” จะเป็นสิ่งที่ช่วยให้การแปลร่ายกรองมีความครบถ้วนทั้งความหมายและ

ฉันทลักษณ์ ดังนั้นขั้นตอนสุดท้ายของระบบจึงออกแบบให้มีส่วนที่ให้ผู้ใช้งานสามารถปรับแต่งร่ายกรองที่แปลเป็นภาษาอังกฤษแล้วได้ตามต้องการโดยไม่บังคับคะแนนความถูกต้องของฉันทลักษณ์ในส่วนนี้ แต่จะใช้เพื่อสะสมคำศัพท์เพื่อประโยชน์ในการแปลครั้งต่อไป และเพื่อสร้างฐานข้อมูลความรู้ให้ระบบสามารถแปลได้มีประสิทธิภาพมากขึ้น

โมเดลการทำงานของกระบวนการที่ 3 จะประกอบด้วย ขั้นตอน และกระบวนการย่อย ดังนี้

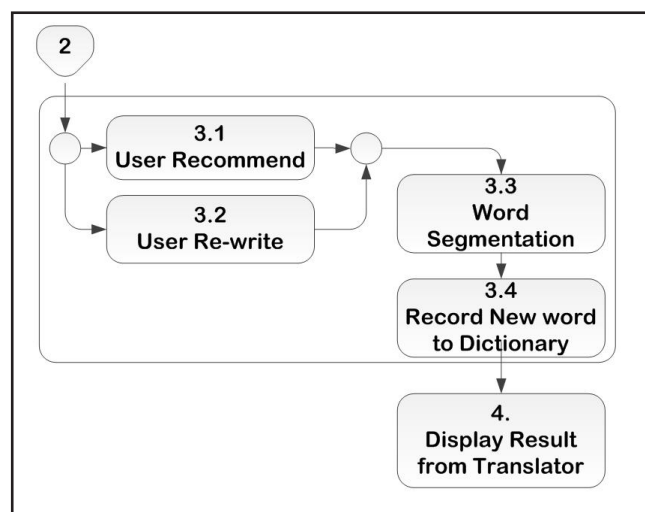
กระบวนการที่ 3.1 ผู้ใช้แนะนำคำศัพท์ ออกแบบให้ผู้ใช้งานแนะนำคำศัพท์ใหม่ หรือ คำศัพท์ที่ผู้ใช้เห็นว่าเหมาะสมได้

กระบวนการที่ 3.2 ผู้ใช้สามารถเขียนกลอนวรรณคดีใหม่ทั้งวรรคได้

กระบวนการที่ 3.3 การตัดคำ จะทำหน้าที่ตัดคำที่ได้ผ่านการแนะนำจากผู้ใช้งาน หรือ เขียนใหม่ทั้งวรรคจากผู้ใช้งาน เพื่อแยกคำศัพท์

กระบวนการที่ 3.4 บันทึกคำศัพท์ที่ได้จากการตัดคำลงพจนานุกรมศัพท์เฉพาะและบันทึกลง Ontology

โมเดลกระบวนการที่ 3 การปรับแต่งและแสดงผล แสดงในภาพที่ 8



ภาพที่ 8 การออกแบบกระบวนการปรับแต่ง และแสดงผล

5. การออกแบบวิธีการวัดผล

เนื่องจากงานวิจัยนี้ต้องการออกแบบขั้นตอนวิธีเพื่อสร้างเครื่องมือในการแปลร่ายกรองจากภาษาไทยไปเป็นภาษาอังกฤษโดยให้คงความหมายและฉันทลักษณ์ที่ถูกต้องให้ได้มากที่สุด ซึ่งเป็นงานใหม่ และไม่สามารถเปรียบเทียบกับประสิทธิภาพและความถูกต้องกับงานวิจัยอื่นได้ ดังนั้นงาน

วิจัยนี้จึงออกแบบให้มีการประเมินโดยวัดเปอร์เซ็นต์ความถูกต้องของฉันทลักษณ์ โดยไม่วัดเรื่องความถูกต้องของเสียงท่ายวรรค แต่ยังคงวัดฉันทลักษณ์ในส่วนของสัมผัส และจำนวนคำ และนับเป็นเปอร์เซ็นต์ความถูกต้อง คือ จากร้อยกรอง 2 บท มีตำแหน่งสัมผัสที่ต้องตรวจสอบ 14 ตำแหน่ง และมีจำนวนพยางค์ที่ต้องตรวจสอบ 8 ตำแหน่ง

5.1 ออกแบบการวัดค่าความถูกต้องต่อสำนวน

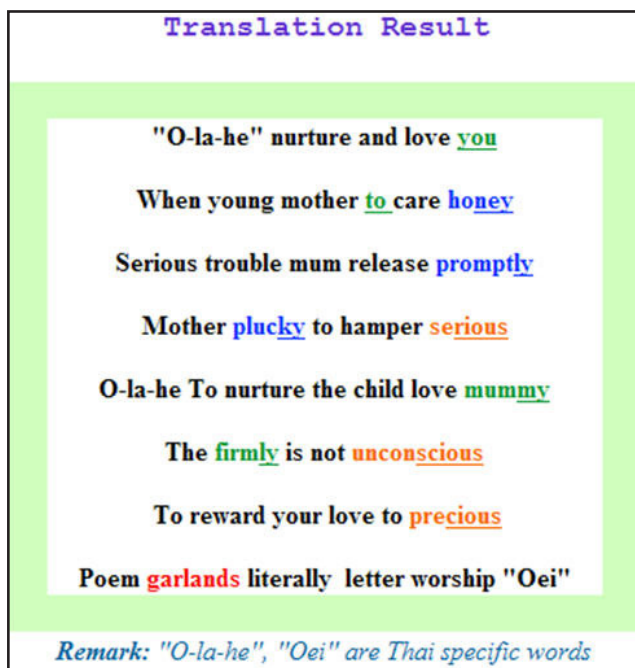
การวัดค่าความถูกต้องต่อสำนวน จะวัดค่าความถูกต้องตามสมการที่ (1) และ วัดค่าความผิดพลาดต่อสำนวนตามสมการที่ (2)

$$P_c = \frac{(R_c + L_c)}{(R_{all} + L_{all})} \times 100\% \quad (1)$$

$$P_w = \frac{(R_w + L_w)}{(R_{all} + L_{all})} \times 100\% \quad (2)$$

เมื่อ P_c หมายถึง เปอร์เซนต์ความถูกต้องต่อสำนวน, P_w หมายถึง เปอร์เซนต์ความผิดพลาดต่อสำนวน, R_c หมายถึง จำนวนสัมผัสที่ถูกต้อง, L_c หมายถึงจำนวนวรรคที่มีพยางค์ถูกต้อง, R_w หมายถึงจำนวนสัมผัสที่ผิดพลาด, L_w หมายถึงจำนวนวรรคที่มีพยางค์ผิดพลาด, R_{all} หมายถึงตำแหน่งสัมผัสที่ตรวจทั้งหมดเท่ากับ 14, L_{all} หมายถึงจำนวนวรรคทั้งหมดเท่ากับ 8

ตัวอย่างการออกแบบหน้าจอรายงานผลจากระบบ แสดงผลลัพธ์ ค่าความถูกต้อง และตำแหน่งที่ผิดพลาด ดังภาพที่ 9



ภาพที่ 8 ส่วนหน้าจอแสดงผลที่ออกแบบ

5.2 ออกแบบวัดค่าความถูกต้องทั้งหมด

การวัดค่าความถูกต้องทั้งหมดในงานวิจัยนี้ วัดค่าความถูกต้องตามสมการที่ (3) และ วัดค่าความผิดพลาดตามสมการที่ (4)

$$T_c = \frac{(PT_c)}{(P_{all})} \times 100\% \quad (3)$$

$$T_w = \frac{(PT_w)}{(P_{all})} \times 100\% \quad (4)$$

เมื่อ T_c หมายถึง เปอร์เซนต์ความถูกต้องทั้งหมด, T_w หมายถึง เปอร์เซนต์ความผิดพลาดทั้งหมด, PT_c หมายถึง จำนวนบทกลอนที่ถูกต้อง, PT_w หมายถึง จำนวนบทกลอนที่ผิดพลาด, P_{all} หมายถึงจำนวนบทกลอนทั้งหมดในการทดสอบ

ทั้งนี้การวัดค่าความถูกต้องที่ออกแบบในงานวิจัยนี้สามารถนำไปใช้เปรียบเทียบกับการประเมินความพึงพอใจและหาความแม่นยำได้ในอนาคต

6. บทสรุปและแนวทางในอนาคต

จากการออกแบบต้นแบบตัวแปลร้อยกรองไทยเป็นภาษาอังกฤษแบบคงฉันทลักษณ์พบว่าปัญหาและอุปสรรคสำคัญในการออกแบบและพัฒนาตัวต้นแบบนี้จำแนกได้เป็น 2 ประเภท คือ ปัญหาที่เกี่ยวข้องกับภาษาไทย และ ปัญหาที่เกี่ยวข้องกับภาษาอังกฤษ ดังนี้

6.1 ปัญหาที่เกี่ยวข้องกับภาษาไทย

1) ปัญหาด้านการตัดคำ เป็นปัญหาที่อาจกล่าวได้ว่ายังคงพบอยู่ทั่วไปในงานวิจัยด้านการประมวลผลภาษาไทย เนื่องจากหากตัดคำผิด ความหมายก็จะผิดเพี้ยนไปด้วย ผู้วิจัยออกแบบให้มีการเพิ่มเงื่อนไขการตรวจสอบวากยสัมพันธ์ของประโยค และกำหนดความเป็นไปได้ในการตัดคำเพิ่มเติมจากโปรแกรมอินเทอร์เน็ตเฟสที่ผู้วิจัยนำมาใช้ และออกแบบให้ผู้ใช้สามารถแนะนำการตัดคำได้

2) คำที่พบในบทกวีนั้นบางคำเป็นคำที่ไม่มีมีความหมายตรงตัว แต่ผู้แต่งมีการปรับคำหรือสลับพยางค์เพื่อให้เกิดการ "รับ" และ "ส่ง" สัมผัสที่ถูกต้อง เช่นคำว่า "โทรมทรุด" ไม่มีความหมายตามพจนานุกรม แต่ที่แท้จริงแล้วมาจากคำว่า "ทรุดโทรม" ซึ่งผู้วิจัยได้ออกแบบให้มีพจนานุกรมศัพท์เฉพาะ และออนไลน์เพื่อใช้ในการค้นหาความหมาย

3) คำในภาษาไทยทำหลายหน้าที่ เช่น คำว่า "หุ่น" เป็น

ได้ทั้งคำนาม หรือ กริยา ทั้งนี้การทำหน้าที่ของคำที่ต่างกัน จะทำให้ความหมายแตกต่างกัน ผู้วิจัยจึงแก้ปัญหาโดย ออกแบบให้มีเงื่อนไขในการตรวจสอบบริบทของคำ

4) คำทับศัพท์ แม้จะมีการออกแบบให้พจนานุกรมศัพท์ เฉพาะเก็บคำทับศัพท์ไว้จำนวนหนึ่ง แต่ในภาษาไทยบาง ครั้งคาดเดายากว่า คำที่พบเป็นคำทับศัพท์หรือไม่ เช่น คำว่า “สัจจะ” เป็นได้ทั้งคำที่ต้องแปล หรือในบทกลอนบาง บท อาจจะหมายถึงคนชื่อ “สัจจะ” ก็ได้

6.2 ปัญหาที่เกี่ยวข้องกับงานภาษาอังกฤษ

1) เสียงในภาษาอังกฤษ เนื่องจากภาษาอังกฤษมีสระ เพียง 5 ตัว และมีพยัญชนะเพียง 26 ตัว การผสมเสียงจึงไม่ หลากหลายเท่าภาษาไทย ดังนั้นโอกาสในการหาคำที่มีความ หมายตรงพบแล้ว แต่เกิดสัมผัสซ้ำหรือ ไม่สัมผัส จึงมีโอกาส สูง

2) ความสัมพันธ์ของคำศัพท์ เนื่องจากเมื่อต้องแปลจาก ภาษาไทยเป็นภาษาอังกฤษให้คงไว้ทั้งความหมายและ ฉันทลักษณ์ บางครั้งไม่สามารถใช้คำที่มีความหมายตรงตัว มาใช้ในบทกลอนได้ ดังนั้นจึงจำเป็นต้องสร้างออนโทโลยี ของคำภาษาอังกฤษโดยอาศัยพื้นฐานของ word net แต่ เชื่อมโยงเพิ่มเติมให้มีความสัมพันธ์สำหรับคำที่ใกล้เคียง หรือความหมายคล้ายกันให้มากขึ้น

3) การแปลคำทับศัพท์ ผู้วิจัยพบว่า เมื่อบทกลอนมีคำที่ จะต้องแปลแบบทับศัพท์จะมีปัญหาในการสื่อสารให้ผู้ ใช้ เข้าใจ ดังนั้นในหน้าอินเทอร์เน็ตเฟสของผู้ใช้ จึงออกแบบให้ ระบบสามารถแสดงหมายเหตุด้านล่างบทกลอนได้ว่าคำใด เป็นคำที่ไม่ต้องแปลความหมาย เนื่องจากทับศัพท์ภาษาไทย

6.3 แนวทางในอนาคต

จากการออกแบบขั้นตอนวิธีที่นำเสนอในบทความนี้ผู้วิจัย กำหนดแนวทางในการพัฒนาตัวต้นแบบ และนำไปทดสอบ กับบทกลอนที่ได้รับความนิยมจากสมาคมกวีร่วมสมัย และบทกลอนที่ได้รับอนุญาตจากเจ้าของสำนวนทั้งสิ้น 500 สำนวน โดยส่วนสำคัญที่จะต้องทำไปควบคู่กันคือการสร้าง พจนานุกรมศัพท์เฉพาะและ ออนโทโลยีของคำศัพท์ให้มีความรู้ที่มีปริมาณที่เพียงพอ พร้อมทั้งสามารถเพิ่มข้อมูล ความรู้ลงในฐานความรู้เหล่านี้จากผู้ใช้ได้ ซึ่งจากต้นแบบ เครื่องมือนี้ จะสามารถพัฒนาต่อยอดให้เครื่องมือมีความ สามารถมากขึ้น แปลกลอนได้แม่นยำถูกต้องตามฉันทลักษณ์ มากขึ้น รวมถึงสามารถพัฒนาสร้างต้นแบบตัวแปลร้อยกรอง

ประเภทอื่นๆได้อีกต่อไป

ร้อยกรองไทยแปลเป็นภาษาอังกฤษแบบคงฉันทลักษณ์ นับเป็นส่วนหนึ่งที่ส่งเสริมการเผยแพร่ความสวยงามของ วรรณศิลป์ไทยให้ได้รับความสนใจยิ่งขึ้นจากคนทั่วโลก

7. เอกสารอ้างอิง

- [1] สมาคมภาษาและหนังสือแห่งประเทศไทย ในพระบรม ราชูปถัมภ์, *อนุสรณ์สุนทรภู่ 200 ปี*. อมรินทร์การพิมพ์. พ.ศ. 2529. ISBN 974-87416-1-3
- [2] โอภาส วงษ์ทวีทรัพย์, นางสาวสิริพร แซ่ลี, นายสมชาย ดิยรางกูร, “การออกแบบภาษาเชิงไวยากรณ์ สำหรับสร้างเครื่องมือตรวจสอบฉันทลักษณ์ คำประพันธ์ไทย,” *5th International Joint Conference on Computer Science and Software Engineering (JCSSE 2008)*, May 2008.
- [3] สัจจาภรณ์ ไวจรรยา, เยาวดี เต็มธนาภักดิ์, “การวัด ประสิทธิภาพของตัวตรวจทานฉันทลักษณ์ และ คุณภาพของกลอนสุภาพ (Performance Evaluation of Thai Poetry (Klon8) Syntax and Quality Analyzer),” *12th National Computer Science and Engineering Conference (NCSEC 2008)*, November 2008.
- [4] L. Balasundaraman, S. Ishwar, S.K. Ravindranath, “Context Free Grammar for Natural Language Constructs an Implementation for Venpa Class of Tamil Poetry”, in *Proceedings of Tamil Internet, India*, 2003.
- [5] Martin Tsan Wong and Andy Hon Wai Chun, “Automatic Haiku Generation Using VSM”, *7th WSEAS Int. Conf. on Applied Computer & Applied Computational Science (ACACOS '08)*, Hangzhou, China, April 6-8, 2008.
- [6] Nathanael Fillmore, *A Romantic Poetry Generation*.
- [7] Lixin Wang, Dan Yang, Junguo Zhu, “A Study of Computer Aided Poem Translation Appreciation,” *Second International Symposium on Knowledge Acquisition and Modeling*, 2009.
- [8] Dmitriy Genzel, Jakob Uszkoreit, Franz Och,



“Poetic Statistical Machine Translation: Rhyme and Meter,” Google, Inc., *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 158–166, USA, 2010.

- [9] ยุทธ โตดิเทพย์ และ สุธีร์ พุ่มกุมาร. *คู่มือเรียนเขียนกลอน*. กรุงเทพฯ: แม่โพสพ, 2548.
- [10] Products | NAIST, “KU Word Cut Demo” [online], Available: <http://naist.cpe.ku.ac.th/wordcut/>