

การแปลภาษาอังกฤษ-ไทย ด้วยคอมพิวเตอร์ โดยอาศัยการจับกระบวนประโยคเข้าคู่

แก้วใจ จันทร์เจริญ¹

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี บางมด ทุ่งครุ กรุงเทพฯ 10140

บทคัดย่อ

การแปลเป็นทั้งศาสตร์และศิลป์ ต้องอาศัยทั้งความรู้ ทักษะ และความชำนาญ ผู้แปลที่ดีจะต้องมีความรู้ทางหลักภาษาหรือโครงสร้างทางไวยากรณ์ของทั้งสองภาษา คือภาษาต้นฉบับ (source language) และภาษาเป้าหมาย (target language) เป็นอย่างดี การศึกษาเพื่อเปรียบเทียบระหว่างโครงสร้างของประโยคในภาษาไทยและภาษาอังกฤษ จึงมีความสำคัญอย่างยิ่งต่อกระบวนการแปลระหว่างภาษาทั้งสอง ทั้งในกระบวนการแปลด้วยมนุษย์ และกระบวนการแปลด้วยเครื่องคอมพิวเตอร์ (machine translation) จากการศึกษาโครงสร้างทางไวยากรณ์ของประโยคภาษาไทยและภาษาอังกฤษพบว่า มีปรากฏการณ์ทางภาษาศาสตร์ที่คล้ายคลึงกันบางประการในโครงสร้างของประโยค เช่น การเรียงลำดับของหน่วยไวยากรณ์ในโครงสร้างประโยคพื้นฐาน (basic sentence pattern) และจากลักษณะดังกล่าวนี้ สามารถนำมาใช้ในการสร้างเทคนิคการจับกระบวนประโยคเข้าคู่ (sentence pattern matching) เพื่อให้ระบบคอมพิวเตอร์สามารถทำการจับคู่เหมือนของกระบวนประโยคในภาษาทั้งสองได้ และนำไปสู่ขั้นตอนการแปลจากประโยคภาษาอังกฤษไปเป็นภาษาไทยได้อย่างถูกต้อง ตามกระบวนการแปลที่มีขั้นตอนเลียนแบบวิธีการแปลด้วยมนุษย์

¹ นักวิจัย ศูนย์ปัญญาประดิษฐ์ สำนักวิจัยและบริการวิทยาศาสตร์และเทคโนโลยี

3 DMUQEDVHG0 DFKIQH 7 UDQVDMRQ

IRU QJDK7KDL

. DHFKDL&KDQFKURHQ ¹

. IQ 0 RQJNXW8 QYHUXW RI 7 HFQRBJ\ 7 KRQEXL Bangmod, RRQJUX%DJNRN

\$ EYWDFW

7 UDQVDMRQIMDSURFHWRI WDMHUIQI WRXJ KWDQGIGHMURP RQHDQI XDJ HRDQRVHIE
 VXEVWVQI Dtext : a VRXLH RUDM IQDMU HDQI XDJ H/ KH-SURFHWRI WDMQDMRQH HP SOLIV
 DFRP EIQDMRQRI DUMWEDQGVIHQWIE P HMRG\$ WDMQDMRUKRXGKDYHDFHUMQFRP SHMQFHRI
 ERWKOQI XDJ HM/ KHFRP SDUDMHWMG EHZ HQ7 KIDQG(QI DMKVHQMCFHMMXFWUHSO) VDFUXFIDO
 URHQWDMQDMQI IQFOXIQI KXP DQWDMQDMRQDQGP DFKIHMDMQDMRQ\$ IVMWVG IQI IQGHSVKVHU
 VHQMQHVMXFWUHWRP HMP IDUQI XVMW SKHQRP HQRIEMKHV QMFWF RUGHRI HFKJ UDP P DMFDO
 XQIQDMHMQFHFRQWVFRQ6 XEMFWHLE 2 EYFMDEHOFDSVUHG7 KIVRIQFICHQMDXUHV
 HS DMGWEHDQHIFVMHMFKQIXHRUHQMFHSDMLOP DFKIQI 7 KHFRP SXMV VMP IV
 GHYCRSHGWEHEDMRP DMKWHHXIYDQMQMCFHSDMLOQIQRVWQI XDJ HV,MMHQMDO
 EURXJ KMERXVQI DMK7 KDMHMQFHFP DFKIHMDMQDMRQV VMP IP DMQI DSURFHWRI KXP DQ
 WDMQDMRQ

¹ GUCIB GI# TUBHCNPNM QPR%QVET+PUNWQI5BCPVHECPF 6CJ PQQJ KCM GUCIB CPF 5GXBU

บทนำ

ในการวิจัยและพัฒนาระบบการแปลภาษาด้วยเครื่องคอมพิวเตอร์ (machine translation system) นั้น มีกลวิธีที่ใช้ในการแปลหลายวิธีด้วยกัน [1] เช่น กลวิธีการแปลโดยตรง (direct translation approach) กลวิธีการแปลแบบใช้การเปลี่ยน (transfer translation approach) กลวิธีการแปลแบบใช้ภาษากลาง (interlingua approach) นอกจากนี้กลวิธีที่เป็นที่รู้จักกันอย่างกว้างขวางเหล่านี้แล้ว ยังได้มีผู้คิดค้นวิธีการแปลอื่นๆ ซึ่งพยายามปรับปรุงให้มีประสิทธิภาพมากยิ่งขึ้น [2] ได้แก่ การแปลโดยใช้ตัวอย่างประโยค (example-based machine translation) การแปลโดยใช้สถิติ (statistic-based machine translation) แม้ในปัจจุบันยังคงมีผู้คิดค้นกลวิธีการแปลภาษาด้วยคอมพิวเตอร์อื่นๆ อีก อาทิเช่น ในโครงการวิจัยการแปลภาษาอังกฤษ-ไทยด้วยคอมพิวเตอร์ โดยอาศัยการจับคู่กระสวนประโยค (pattern-based machine translation : PBMT) ดังที่จะได้นำเสนอในบทความนี้

การแปลด้วยกลวิธีนี้อาศัยการพิจารณาเปรียบเทียบระหว่างโครงสร้างประโยคของภาษาต้นทาง (source language) กับโครงสร้างประโยคภาษาเป้าหมาย (target language) เพื่อการจับรูปแบบโครงสร้างประโยคหรือกระสวนประโยคพื้นฐาน (basic sentence pattern) ที่มีความคล้ายคลึงเข้าคู่กัน (matching) โดยใช้เกณฑ์ทั้งทางด้านวากยสัมพันธ์หรือหน้าที่ของคำ และทางอรรถศาสตร์หรือความสัมพันธ์ทางความหมายเป็นตัวบ่งชี้ว่า โครงสร้างประโยคใดคล้ายคลึงกันหรือต่างกัน แล้วจึงสร้างตารางเพื่อการจับเข้าคู่ (mapping table) จากนั้นจึงแปลคำศัพท์จากภาษาอังกฤษเป็นภาษาไทยชนิดคำต่อคำ เนื่องจากคำศัพท์บางคำในภาษาอังกฤษสามารถแปลเป็นคำไทยได้หลายคำ ประโยคภาษาไทยที่เกิดจากการแปลด้วยกลวิธีนี้จึงมีได้หลายประโยค ดังนั้นวิธีการแก้ความกำกวมทางความหมาย (word sense disambiguation) ด้วยสถิติ (statistic method) และการศึกษาคำเกิดร่วม (word collocation) จึงได้ถูกนำมาใช้เพื่อแก้ปัญหาดังกล่าว

กลวิธีที่นำเสนอในบทความนี้อาจเป็นเพียงทางเลือกใหม่อีกทางหนึ่งในบรรดาแนวทางต่างๆ ที่มีผู้คิดค้นขึ้นและพยายามพัฒนาอย่างไม่หยุดยั้ง เพื่อให้ได้มาซึ่งระบบแปลภาษาที่มีประสิทธิภาพสามารถใช้งานได้จริงและสะดวกต่อผู้ใช้ ข้อดีของระบบการแปลโดยอาศัยการจับคู่กระสวนประโยค (pattern-based machine translation) นี้ คือการที่ระบบสามารถเรียนรู้โครงสร้างประโยคใหม่ๆ ได้เพิ่มขึ้นเรื่อยๆ กล่าวคือ หากระบบสามารถจับโครงสร้างประโยคภาษาต้นทางเข้าคู่กับโครงสร้างประโยคภาษาเป้าหมาย ดังที่ปรากฏในตารางการจับเข้าคู่ได้ ระบบจะทำการแปลตามขั้นตอน แต่หากโครงสร้างของประโยคภาษาต้นทางมีความแตกต่างไปจากโครงสร้างประโยคพื้นฐานในตารางการจับเข้าคู่ โครงสร้างนั้นๆ จะถูกจัดเก็บไว้ เพื่อนำมาเพิ่มเติมเข้าไปในตารางในภายหลัง ในที่สุดระบบจะสามารถทำการแปลประโยคภาษาอังกฤษที่มีโครงสร้างต่างไปจากโครงสร้างประโยคพื้นฐาน หรือโครงสร้างประโยคที่ซับซ้อนขึ้นได้ อย่างไรก็ตามการพัฒนากระบวนการแปล PBMT นี้ยังอยู่ในระยะเริ่มต้น ซึ่งมีข้อบกพร่องที่ต้องแก้ไขอีกเป็นอันมาก ทั้งนี้คณะผู้วิจัยจะได้ทำการศึกษาค้นคว้า และแก้ไขโดยลำดับ

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและพัฒนาระบบการแปลภาษาด้วยคอมพิวเตอร์ให้มีประสิทธิภาพมากยิ่งขึ้น
2. เพื่อหาแนวทางใหม่ในการพัฒนาระบบการแปลภาษาด้วยคอมพิวเตอร์
3. เพื่อสร้างระบบที่สามารถนำมาใช้ในการแปลภาษาได้จริงและเป็นระบบที่สามารถทำการพัฒนาให้มีความฉลาดยิ่งขึ้น

ขอบเขตการวิจัย

1. ในการวิจัยครั้งนี้ผู้วิจัยมุ่งศึกษากลวิธีการแปลประโยคภาษาอังกฤษเป็นภาษาไทย โดยอาศัยการสร้างตารางจับคู่ ซึ่งพิจารณาจากความคล้ายคลึงของโครงสร้างประโยคทั้งทางด้านวากยสัมพันธ์และความสัมพันธ์ทางความหมายหรือ การกสัมพันธ์
2. ผู้วิจัยทำการศึกษาเฉพาะการแปลโครงสร้างประโยคพื้นฐานเท่านั้น โดยยังมีได้ศึกษาโครงสร้างประโยครูปแบบอื่นๆ ที่มีความซับซ้อน เช่น ประโยคความรวม ประโยคความซ้อน หรือประโยคที่มีลักษณะพิเศษ เช่น ประโยคที่มีการละคำ หรือการเปลี่ยนลำดับของคำ

ลักษณะโครงสร้างของประโยคพื้นฐานภาษาอังกฤษและภาษาไทย

จากการทบทวนวรรณกรรมที่มีผู้ศึกษาระบบไวยากรณ์ไทย [3,4,5] ไว้ อาจารย์รวบรวมคำนิยามของประโยคในภาษาไทยได้ว่า “ประโยคคือ ถ้อยคำที่มีเนื้อความบริบูรณ์ ประกอบด้วยส่วนสำคัญสองส่วนซึ่งแสดงความคิดที่เกี่ยวข้องสัมพันธ์กัน อันได้แก่ ภาคประธาน หรือหน่วยนาม และภาคแสดงหรือหน่วยกริยา” และหากพิจารณาประโยคในภาษาอังกฤษตามที่ผู้ให้คำนิยามไว้ [6] พบว่ามีลักษณะใกล้เคียงกัน โครงสร้างของประโยคทั้งในภาษาไทยและภาษาอังกฤษแบ่งออกได้เป็น 3 ชนิดคือ

1. ประโยคความเดียว (simple sentence) ประกอบด้วยหน่วยนามเป็นภาคประธาน และหน่วยกริยาเป็นภาคแสดง
2. ประโยคความซ้อน (complex sentence) เกิดจากการนำประโยคมาขยายหน่วยนาม หรือหน่วยกริยาในประโยคความเดียว เพื่อให้ประโยคมีความสลับซับซ้อนและมีรูปแบบมากยิ่งขึ้น
3. ประโยคความรวม (compound sentence) เกิดจากการนำประโยคความเดียวตั้งแต่ 2 ประโยคขึ้นไป มาเชื่อมเข้าด้วยกันด้วยคำเชื่อม

เมื่อลองพิจารณาโครงสร้างประโยคภาษาไทยและภาษาอังกฤษในรายละเอียด พบว่าโครงสร้างประโยคพื้นฐานมีลักษณะใกล้เคียงกัน ซึ่งอาจเทียบเคียงได้ดังนี้

ตารางที่ 1 : การเทียบเคียงโครงสร้างประโยคพื้นฐานภาษาไทยและภาษาอังกฤษ

โครงสร้างประโยคภาษาไทย	โครงสร้างประโยคอังกฤษ
หน่วยนาม+หน่วยกริยาแสดงอาการ	np+verb
หน่วยนาม1+หน่วยกริยาแสดงอาการ+หน่วยนาม2	np+verb+np2
หน่วยนาม+หน่วยกริยาเปรียบเทียบ+หน่วยขยายนาม (คำนาม)	np+be+complement
หน่วยนาม+หน่วยกริยา (แสดงสภาพ/แสดงความรู้สึก)	np+be+adj
หน่วยนาม+หน่วยกริยาแสดงอาการ+หน่วยขยายกริยา/หน่วยนามบอกสถานที่	np+be+adv/pp
หน่วยนาม1+หน่วยกริยาแสดงอาการ+หน่วยนาม2+หน่วยขยาย (กรรมตรง)	np1+verb+np2+object complement
หน่วยนาม1+หน่วยกริยาแสดงอาการ+หน่วยนาม2+หน่วยนาม3	np1+verb+np3+np2
	It+be+complement
	There+inverted clause

ตัวอย่าง :

ประโยคภาษาไทย

1. หมาเห่า
2. หมาไล่แมว
3. หมาเป็นสัตว์
4. หมาน่ารัก
5. จอห์นอยู่ที่นี้
6. หมาทำให้จอห์นสุขใจ
7. พวกเขาเลือกจอห์นเป็นประธานาธิบดี
8. เขาให้เงินน้องสาว
9. -
10. -

ประโยคภาษาอังกฤษ

- ' RJV EDUN
' RJVFKDMHFDW
' RJVDUHDQP DO
' RJVDUHQFH
-RKQV/KHUH
' RJVP DNH- RKQKDSS\
7KH\ VHGFVWG-RKQSUHMGHQW
+HJDYHKLVVWVUP RQH
MUDIQJ
7KHUHLVDP DQRXVWGH

จากตารางเทียบเคียงโครงสร้างของประโยคพื้นฐานข้างต้น จะเห็นได้ว่านอกจากโครงสร้างของประโยคในภาษาทั้งสองจะมีลักษณะคล้ายคลึงกันแล้ว ในด้านการเรียงลำดับของคำยังคล้ายกันอีกด้วย คือมีการเรียงลำดับคำในโครงสร้างประโยคเป็นแบบ ประธาน-กริยา-กรรม [7] ความคล้ายคลึงกันทั้งสองข้อนี้เองที่คณะผู้วิจัย ได้นำมาใช้ประโยชน์ในการจับคู่โครงสร้างประโยคหรือกระบวนประโยคแบบต่างๆ (ซึ่งจะได้กล่าวถึงในตอนถัดไป)

อย่างไรก็ตามเมื่อพิจารณาหน่วยย่อยต่าง ๆ ที่เป็นองค์ประกอบในโครงสร้างประโยค ทั้งหน่วยนามหรือนามวลี และหน่วยกริยาหรือกริยาวลี พบว่ามีความแตกต่างกันทางโครงสร้างกล่าวคือ ในส่วนของนามวลี หน่วยขยายของนามวลีในภาษาไทย อันได้แก่ คำนาม คำบอกบุรุษ คำกริยา คำบอกจำนวน คำบอกลำดับ คำบ่งชี้ คำบอกความสำคัญ คำไม่ชี้เฉพาะ และอนุประโยค จะอยู่ในตำแหน่งหลังคำนามที่เป็นหน่วยหลัก แต่ในภาษาอังกฤษ [8] นั้น หน่วยขยายอาจวางอยู่ทั้งที่หน้าคำนามและหลังคำนาม

ตัวอย่าง :

เขามีบ้านหลังใหญ่มหึมา

เด็ก 2 คนตกน้ำตาย

งานชิ้นนี้เกือบเสร็จแล้ว

เด็กคนที่ฉันเคยช่วยเหลือตายเสียแล้ว

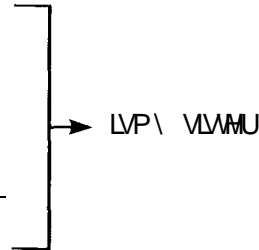
The ju

7KH pretty JUUD

7KH pretty JUUD IQMHFRUQU

7KH SUHWV JUUD ZKR VWVQGGQ IQMHFRUQU

6KH



ในส่วนของหน่วยกริยาหรือกริยาวลีพบว่า หน่วยขยายในโครงสร้างกริยาวลีของภาษาไทย อาจอยู่หน้าหรือหลังคำกริยาหลักและอาจไม่ได้ยู่ชิดกับหน่วยหลัก สามารถมีคำอื่นมาคั่นได้ นอกจากนี้กลวิธีการแสดง “กาล (tense)” และ “การณัลักษณะ (aspect)” ในกริยาวลีของภาษาไทยนั้น ใช้การเติมคำกริยาช่วย การซ้ำกริยา และการเรียงกริยา [9] ขณะที่ในภาษาอังกฤษใช้การเติมคำกริยาช่วยและการเติมปัจจัย (suffix) ที่ท้ายคำกริยาหลัก

ตัวอย่าง :

จอห์นกำลังทำงานอยู่

John was working.

จอห์นมาถึงแล้ว

John has arrived.

ความแตกต่างเหล่านี้ส่งผลกระทบต่อการแปลด้วยวิธีการจับคู่โครงสร้างประโยค ทั้งนี้เนื่องจากการจับคู่โครงสร้างที่เหมือนกันตามตารางการจับคู่แล้วทำการแปลชนิดคำต่อคำสามารถดำเนินการได้ค่อนข้างสมบูรณ์ เมื่อโครงสร้างประโยคที่ต้องการแปลมีเพียงหน่วยนามหรือหน่วยกริยาปรากฏโดดๆ โดยไม่มีหน่วยขยายหรือไม่มีการแสดง “กาล” และ “การณัลักษณะ” แต่หากโครงสร้างประโยคภาษาอังกฤษหรือประโยคภาษาต้นทางมีหน่วยขยายใดๆ ซึ่งมีความแตกต่าง การใช้กฎเฉพาะเพื่อช่วยแปลงโครงสร้างหน่วยนามและหน่วยกริยาจากภาษาอังกฤษให้เป็นภาษาไทยที่ถูกต้อง จึงเป็นสิ่งจำเป็น

ตัวอย่างกฎโครงสร้างวลีสำหรับภาษาอังกฤษ [8]

- ← NP Aux VP
- ← H₁GM 33
- ← 1 6 ¶
- 93 ← 13 33
- ← EH3UHG
- ← 3UH13
- ← 7HQV0 RG (Perf) URJ
- UHG ← GY13 \$GM

ตัวอย่างกฎโครงสร้างวลีสำหรับภาษาไทย

- ← 1 3 9
- ← 1 \$GMW' H₁B
- ← 1 6 ¶
- ← 13 33
- ← UH13

ขั้นตอนการแปลประโยคภาษาอังกฤษ-ภาษาไทยด้วยคอมพิวเตอร์ โดยวิธีการจับคู่กระสวนประโยค (pattern-based machine translation)

ความพยายามในการพัฒนาระบบแปลภาษาด้วยคอมพิวเตอร์ในต่างประเทศมีมาอย่างต่อเนื่อง มีผู้ทดลองคิดค้นและหาแนวทางการพัฒนากลวิธีการแปลใหม่ๆ เพื่อให้ระบบมีประสิทธิภาพสูงขึ้น โดยนำข้อดีของกลวิธีเดิมๆ มาดัดแปลงและประยุกต์เข้าด้วยกันเพื่อสร้างเป็นกลวิธีใหม่ กลวิธีการแปลที่น่าสนใจกลวิธีหนึ่ง ได้แก่ การแปลโดยอาศัยตัวอย่างประโยค (example-based machine translation) [10] ซึ่งมีหลักในการแปลโดยการเก็บตัวอย่างข้อความคู่ภาษาที่แปลแล้วจำนวนมากไว้ในคลังข้อมูล จากนั้นจึงทำการค้นคืนตัวอย่างการแปลโดยการเทียบแบบ (analogy) มาใช้เป็นแบบในการแปลประโยคอินพุต ยังมีกลวิธีการแปลอีกหนึ่งกลวิธีที่น่าสนใจกลวิธีการแปลโดยอาศัยตัวอย่างประโยคไปประยุกต์ใช้ นั่นคือการแปลโดยการผสมผสาน หรือที่เรียกว่า กลวิธีผสมผสาน (hybrid translation method) [11] หลักการสำคัญของกลวิธีนี้ได้แก่ การนำกลวิธีการแปลโดยอาศัยตัวอย่างประโยคมารวมเข้ากับกลวิธีการแปลโดยอาศัยกฎ (rule-based method) กล่าวคือ เมื่อทำการค้นคืนตัวอย่างประโยคจากคลังข้อความแปลคู่ภาษาแล้ว จึงตัดให้เหลือประโยคตัวอย่างที่ใกล้เคียงกับประโยคอินพุตมากที่สุด จากนั้นจึงแปลตามประโยคตัวอย่าง หากพบว่ามีส่วนใดของประโยคต้นทางแตกต่างจากประโยคตัวอย่าง จะเขียนกฎเฉพาะ หรือสร้าง Rule-based Module เพื่อทำการแปลส่วนที่ต่างนั้น

ในงานวิจัยขั้นนี้ได้พยายามดัดแปลงกลวิธีต่างๆ ที่กล่าวมานี้มาประยุกต์ใช้ ดังมีขั้นตอนสำคัญ 3 ขั้นตอนคือ

1. ขั้นตอนการวิเคราะห์ (input analysis)

ประโยคภาษาต้นทางจะถูกวิเคราะห์ด้วยกฎโครงสร้างวลี (phrase structure rule) แบบวิธีจากล่างขึ้นบน (bottom-up parsing technique) หากทำการวิเคราะห์ผ่านจะได้โครงสร้างประโยคอินพุตที่อยู่ในรูปของสัญลักษณ์หน่วย NP, VP, PP เป็นต้น แล้วจึงนำลำดับของสัญลักษณ์ (string of non-terminal symbols) เหล่านี้ไปเทียบเคียงกับตารางการจับเข้าคู่ในขั้นตอนต่อไป

2. ขั้นตอนการจับโครงสร้างหรือกระสวนประโยคเข้าคู่ (pattern mapping)

โครงสร้างของประโยคอินพุตที่อยู่ในรูปของลำดับของโครงสร้างวลีต่างๆ จะถูกนำไปเทียบกับโครงสร้างหรือกระสวนประโยคภาษาอังกฤษในตารางการจับเข้าคู่ที่สร้างขึ้น หากจับเข้าคู่ได้ระบบจะดึงคู่โครงสร้างประโยคภาษาไทยที่อยู่ในตารางออกมาด้วย เพื่อทำการเรียงลำดับของ NP, VP, PP... ของประโยคอินพุตเสียใหม่ตามโครงสร้างประโยคภาษาไทย แล้วจึงค่อยส่งไปยังขั้นตอนต่อไป แต่หากโครงสร้างของประโยคอินพุตไม่สามารถจับเข้าคู่ตามตารางได้ โครงสร้างนั้นจะถูกเก็บไว้เพื่อนำไปสร้างตารางการจับเข้าคู่ตารางใหม่ในภายหลัง

ตัวอย่าง : ตารางการจับเข้าคู่โครงสร้างประโยค

โครงสร้างประโยคภาษาอังกฤษ	โครงสร้างประโยคภาษาไทย
1. NP % \$' -	1. NP ADJ (STATIVE VERB)
2. NP % \$' 9	2. NP % \$' 9
3. NP1 % 13	3. NP1 % 13
4. NP	4. NP
5. NP 33	5. NP 33
6. NP1 13	6. NP1 13
7. NP1 13 13	7. NP1 13 13

ในส่วนของโครงสร้างนามวลีและกริยาวลี จะมีตารางย่อยเพื่อการแปลงจากภาษาอังกฤษเป็นภาษาไทย เช่น

(QILMKI RKQ3 KUDH	7KDLI RKQ3 KUDH
DI 3 → 3 5 2	13 → 3 5 2
E 13 → '(13	13 → 1 & \$ 66 ' (7
F1 3 32 66 352 1 13	13 → 1 32 66 352 1
GI 3 \$ 57 13	NP → NP
H 3 → ADJ 1	13 → 1 & \$ 66 \$ ' -
I 1 3 1 3 3 3	1 3 1 3 3 3
g) NP → N	NP → N
h) NP → N+	13 → (reverse)N+

3. ขั้นตอนการแปลแบบคำต่อคำ (word by word translation)

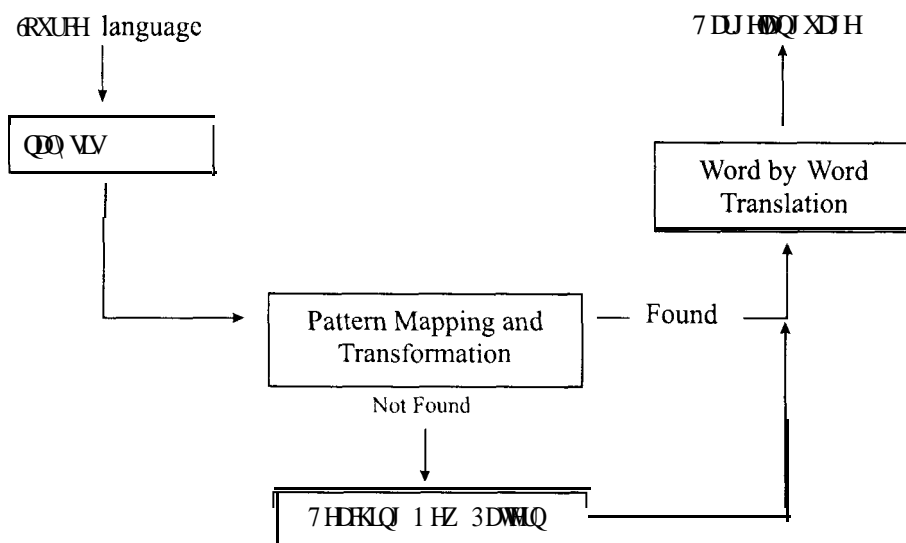
ในขั้นตอนนี้คำศัพท์ภาษาไทยจากพจนานุกรมคู่ภาษาจะถูกนำมาแทนที่คำศัพท์ภาษาอังกฤษ แต่เนื่องจากคำศัพท์ภาษาอังกฤษหนึ่งคำสามารถแปลเป็นคำศัพท์ภาษาไทยได้หลายคำ ประโยคภาษาไทยที่ได้จากการแปลในขั้นตอนนี้ จึงมีหลายประโยค ในขั้นแรกคณะผู้วิจัยจึงได้คิดวิธีการแก้ความกำกวมทางความหมายของคำศัพท์ภาษาไทยด้วยการใส่สถิติให้แก่คำศัพท์แต่ละคำ ตามการใช้คำนั้นๆ ในบริบทที่พบมากที่สุด แต่ยังไม่สามารถลดความกำกวมได้มากเท่าที่ต้องการในทุกๆ ประโยค คณะผู้วิจัยจึงพยายามหาวิธีการแก้ความกำกวมทางความหมายของคำ (word sense disambiguation) ด้วยการศึกษาคำเกิดร่วม (word collocation) ซึ่งขณะนี้อยู่ในระหว่างการทดลองพัฒนาโปรแกรม

การแก้ความกำกวมทางความหมายของคำ (word sense disambiguation) คือ การระบุหรือการให้ความหมายที่ถูกต้องของคำศัพท์ เพื่อไม่ให้เกิดการตีความที่ผิดเพี้ยนไป การแก้ความกำกวมทางความหมายมีประโยชน์อย่างยิ่งต่องานวิจัยด้านการประมวลผลภาษาธรรมชาติ ซึ่งรวมทั้งการแปลภาษาด้วยคอมพิวเตอร์ด้วย หลักการการพัฒนาอัลกอริทึมเพื่อแก้ความกำกวมทางความหมายคือ การใช้ความหมายของคำข้างเคียงหรือคำที่เกิดร่วม เป็นตัวระบุหรือช่วยตัดสินความหมาย (sense) ที่ถูกต้องของคำเป้าหมาย (target word) จากงานวิจัยของ Tanapong Potipiti and Surapant Meknavin [12] ได้นำหลักการดังกล่าวมาใช้ โดยใช้ในการแก้ความกำกวมทางความหมายของคำศัพท์ภาษาไทย ซึ่งมีความหมาย 2 ความหมาย จากประโยคภาษาไทย 5,000 ประโยค ผลการวิจัยพบว่าวิธีนี้ประสบผลสำเร็จถึง 93 เปอร์เซ็นต์

ตัวอย่าง : การแยกความหมาย (sense) ของคำว่า “ เกาะ ” โดยใช้ความหมายของคำเกิดร่วม

6HQHVDOG	6HQH\$ WFK
อินโดนีเซีย -QGRQML	บิน - fly
ทะเล -VD	รัง -QH
อ่าว -ED	นก -HU
หาด -VRUH	ต้นไม้ -WH
เรือ -VL	จับ -FWK
สำรวจ -GEMR&HU	แมลง -IQMF

จากขั้นตอนการแปลโดยการจับคู่โครงสร้างหรือกระสวนประโยคตามที่ได้กล่าวมานี้ อาจแสดงแบบจำลองของระบบได้ดังนี้



รูปที่ 1 แบบจำลอง PBMT

7. สรุปผลการวิจัย

กระบวนการแปลภาษาด้วยเครื่องคอมพิวเตอร์จากภาษาอังกฤษเป็นภาษาไทย ด้วยวิธีการเปรียบเทียบกระสวนประโยค และการจับกระสวนประโยคเข้าคู่ (pattern matching) นี้ สามารถที่จะทำการแปลประโยคต่างๆ ในภาษาอังกฤษเป็นภาษาไทยได้ด้วยหลักการสำคัญๆ คือ การสร้างตารางการจับเข้าคู่กระสวนประโยค (matching table) ตามลักษณะโครงสร้างประโยคหรือกระสวนประโยคนั้นๆ และยังได้สร้างตารางย่อยสำหรับการแปลงโครงสร้างวลีภาษาอังกฤษให้เป็นภาษาไทยอย่างถูกต้อง ในกรณีการแปลประโยคที่คำศัพท์ภาษาอังกฤษมีคำแปลภาษาไทยแบบหนึ่งต่อหนึ่งนั้น ระบบสามารถแปลได้อย่างถูกต้องถึง 90% แต่ในกรณีที่คำศัพท์ภาษาอังกฤษในประโยคต้นทาง

มีคำแปลภาษาไทยมากกว่าหนึ่งคำ ทำให้ระบบแปลประโยคภาษาเป้าหมายได้มากกว่าหนึ่งประโยค จึงได้มีการนำสถิติการเกิดหรือความเป็นไปได้ของคำศัพท์ในบริบทนั้นๆ เข้ามาช่วยในการเลือก เพื่อให้ได้คำในภาษาเป้าหมายที่มีความหมายตรงกันหรือใกล้เคียงที่สุดกับคำในภาษาต้นทาง

ตัวอย่าง :

1. That is a good idea.	→	สิ่งนั้นเป็นความคิดดี	0.72500
		สิ่งนั้นเป็นความเห็นดี	0.65000
		สิ่งนั้นเป็นความคิดถูก	0.65000
RQRWHDQIGRW	→	อย่าเป็นคนปัญญาอ่อน	1.000000
		อย่าอยู่คนปัญญาอ่อน	0.920000
		อย่าคือคนปัญญาอ่อน	0.920000
EURMHJUQGV/DNQIH	→	พี่ชายของฉันบดมีด	RRR
		พี่ชายของฉันลับมีด	RRR
		พี่ชายของฉันฝนมีด	RRR

การใช้วิธีการทางสถิติเข้ามาช่วยในการเลือกคำศัพท์สามารถลดความกำกวมทางความหมายได้ถึง 50-60% อย่างไรก็ตามในกรณีที่คำศัพท์นั้นๆ มีสถิติการเกิดในบริบทในอัตราที่เท่ากัน ดังเช่นประโยคที่ 3 วิธีการแก้ความกำกวมด้วยสถิติอย่างเดียวจึงไม่เพียงพอ ในแผนงานขั้นต่อไป จะได้ทำการแก้ความกำกวมด้วยหลักการการศึกษาคำเกิดร่วมเพื่อแยกแยะความหมาย (sense) ที่ชัดเจนของคำที่ต้องการ หรือการใช้แนวคิด “One sense per collocation constraint” [12] หรือหนึ่งความหมายจากลักษณะบังคับของคำเกิดร่วม เพื่อให้ได้ประโยคภาษาไทยที่ถูกต้องทั้งโครงสร้างทางวากยสัมพันธ์และความหมาย

8. เอกสารอ้างอิง

1. นิตยา กาญจนะวรรณ, 2534, *การแปลภาษาด้วยเครื่อง : ทฤษฎีและวิธีการ*. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยรามคำแหง
2. วิรัช ศรเลิศล้ำวานิช และ สุรพันธ์ เมฆนาวิน, 2537, *แนวโน้มการวิจัยระบบเครื่องแปลภาษา*. สาร NECTEC ปีที่ 2 กรกฎาคม-กันยายน หน้า 42-48
3. นววรรณ พันธุเมธา, 2527, *ไวยากรณ์ไทย*. กรุงเทพฯ : รุ่งเรืองสาส์นการพิมพ์ หน้า 135-155.
4. วิจินตน์ ภาณุพงศ์, 2530, *โครงสร้างของภาษาไทย*. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยรามคำแหง

5. มหาวิทยาลัยสุโขทัยธรรมาราช, 2531, ภาษาไทย 3 : หน่วยที่ 7-15. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาราช

XLN5 HND) TCOOCTQH%QPWDRQICJ ' PINUJ (WH Longman
* URXS/ WSS

Fromkin, DQG5 REHSW #P +PVIQFVEWQP W/ DQ XDJ H) (RUCD/ KH
Dryden UHSS

WQ Natural Language 7 PFGIUWPHH California: The %HVP IQ&XP P IQ V
3 XEOWIQ &RP SDQ Inc.pp.

KHNDQWQD. Serial HE %QUMWIKPIP6JCK 3K' GWHVWNRQ8 QYHMW
RI 0 IEKLDQ

DJ DR0 #) UP HZ RUN QC 0 HKDQLDQ UDQWDRQEHZ HQ- DSDQHMI CPF
' PINUJ E\ \$ QDBJ\ 3 UQFISQIQ\$ (WRP DQG5 %DQHMG\$ UWIFLDQ
+XP DQ QMQLHQFH

KLD6 %RQG) DQG7 DNDKMK< # * [DIF 5 XHDQ([DP SØEDMG
0 HMRG HQT0 LKLC/ UDQWDRQ, QSURFHGQI VRI 7 KHI DMLDQDQ XDJ HB URFHMQI
3 DHULF5 IP 6\ P SRMXP ' HFP EHSS

WBMZ DQG0 HNDMQ6 FUWRCMUCF %QIRWDCUCF 9 QIF 5 CUG
Disambiguation. , Q SURFHGQI VRI WH 6FIHQMF &RQIHUQH - 2FWHU
+DQR9 IEMQP