

บทที่ 1

บทนำ

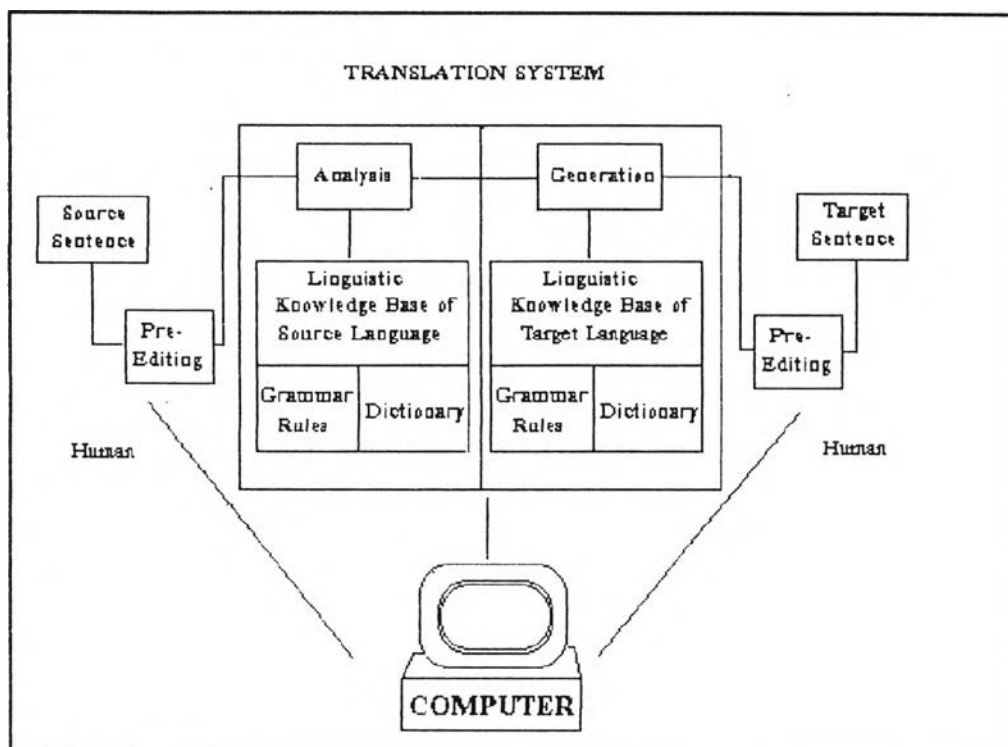


1.1 ความสำคัญและความเป็นมาของปัญหา

การแปลภาษาจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่งเป็นสิ่งจำเป็นและมีความสำคัญมากในหลายๆ ด้าน ตัวอย่างเช่น ในด้านการกระจายเทคโนโลยี การแปลภาษาทำให้นักวิจัยสามารถอ่านสิ่งตีพิมพ์ในภาษาต่างประเทศ เพื่อนำมาใช้ประโยชน์ในการวิจัย ในด้านธุรกิจ นักธุรกิจสามารถใช้ข้อมูลข่าวสารที่เป็นภาษาต่างประเทศมาเป็นส่วนประกอบในการตัดสินใจ ในด้านการสื่อสารและการท่องเที่ยว นักท่องเที่ยวสามารถรู้ถึงขนบธรรมเนียมประเพณีของชาติที่สนใจ ทำให้เกิดการแลกเปลี่ยนวัฒนธรรมและเกิดความเข้าใจอันดีระหว่างคนของทั้งสองภาษานั้น

ในการแปลภาษาหนึ่งไปเป็นอีกภาษาหนึ่งด้วยคอมพิวเตอร์นั้นเป็นการเลียนแบบกระบวนการแปลภาษาของมนุษย์ [1] คือ มนุษย์จะต้องอ่านภาษาต้นแบบ และทำความเข้าใจกับภาษาต้นแบบนั้น แล้วจึงแปลมาเป็นภาษาเป้าหมายตามความคิดที่มนุษย์เข้าใจ ซึ่งผู้แปลนั้นจะต้องเข้าใจ หรือมีความรู้ ทั้งภาษาต้นแบบและภาษาเป้าหมาย ซึ่งมนุษย์จะใช้ความรู้ที่จดจำไว้ในสมองเป็นเครื่องมือในการแปล ส่วนระบบคอมพิวเตอร์จะแปลภาษาได้ ต้องอาศัยกระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing) และฐานความรู้ทางภาษาศาสตร์ (Linguistic Knowledge Base)

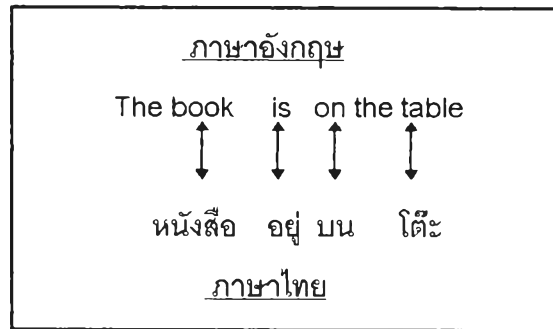
ดังนั้น การแปลภาษาด้วยระบบคอมพิวเตอร์จึงต้องเริ่มจากขบวนการวิเคราะห์ประโยคของภาษาต้นแบบ (Source Sentence) เพื่อหาข้อมูลทางด้านโครงสร้าง (Syntax) และความหมาย (Semantic) โดยอาศัยฐานความรู้ทางภาษาศาสตร์ของภาษาต้นแบบนั้น ได้แก่ พจนานุกรมและกฎไวยากรณ์ต่าง ๆ ดังแสดงในรูปที่ 1



รูปที่ 1 โครงสร้างของระบบการแปลภาษาด้วยคอมพิวเตอร์ทั่วไป

ในการแปล ภาษาด้วยระบบคอมพิวเตอร์ทั่วไป ในบางครั้งอาจมีความจำเป็นต้องจัดการกับประโยคต้นแบบ (Pre-editing) ก่อนที่จะป้อนเข้าสู่กระบวนการวิเคราะห์ เช่น การตรวจสอบตัวสะกด (Spelling Check) หรือการตัดประโยค เป็นต้น ในทำนองเดียวกัน อาจมีความจำเป็นต้องจัดการกับผลลัพธ์ประโยคเป้าหมายที่สังเคราะห์ได้ (Post-editing) ให้มีความถูกต้องยิ่งขึ้น เช่น มีความเป็นธรรมชาติในทางความหมายของภาษามากขึ้น เป็นต้น

หากพิจารณาการแปลภาษาจากภาษาหนึ่งไปยังอีกภาษาหนึ่งนั้น บางครั้งจะดูเหมือนว่าสามารถแปลในลักษณะหนึ่งต่อหนึ่งได้ [2] ดังแสดงในรูปที่ 2



รูปที่ 2 ตัวอย่างการแปลจากภาษาอังกฤษเป็นภาษาไทย

หากดูโครงสร้างของภาษาแล้วจะพบว่า การแปลภาษาจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่ง จะต้องเกี่ยวกับความรู้ทางภาษาศาสตร์ในหลายระดับ ทั้งระดับรูปคำ ระดับวากยสัมพันธ์ (syntax) และระดับความหมาย เช่น ถ้าหากต้องการแปลข้อความว่า

The coach lost a set ให้เป็นความหมายภาษาไทย เราจะพบปัญหามากมาย เช่น เมื่อนำคำหลักมาดูในรูปวากยสัมพันธ์จะพบว่า

คำว่า coach เป็นได้ทั้ง คำนาม และกริยา

คำว่า lost เป็นได้ทั้ง กริยา และวิเศษณ์

คำว่า set เป็นได้ทั้ง คำนาม กริยา และวิเศษณ์

ดังนั้นความเป็นไปได้ของประโยคทางวากยสัมพันธ์เป็นไปได้ถึง $2 \times 2 \times 3$ หรือ 12 แบบ ซึ่งก็จะต้องแยกแยะทางวากยสัมพันธ์ให้ถูกต้องเสียก่อน ถึงแม้ว่าจะดูทางด้านวากยสัมพันธ์แล้วก็ตามก็ยังไม่พอเพียงในการที่จะแปลความหมายได้ เพราะถ้าลองเปิดพจนานุกรมเพื่อดูความหมาย จะพบว่า

คำว่า coach มีความหมายได้ 6 ความหมาย

คำว่า lost มีความหมายได้ 10 ความหมาย

คำว่า set มีความหมายได้ 34 ความหมาย

เมื่อรวมกันจะมีความเป็นไปได้ถึง $6 \times 10 \times 34$ หรือ 2,040 รูปแบบ ดังนั้น ปัญหาของการแปลภาษาด้วยคอมพิวเตอร์ ก็คือ การทำให้ระบบคอมพิวเตอร์เข้าใจโครงสร้างและความหมายของประโยคหรือข้อความที่จะแปลได้เหมือนมนุษย์ [1]

ปัจจุบันคอมพิวเตอร์มีขีดความสามารถในการประมวลผลได้เร็วกว่าเดิมมาก และสามารถเก็บข้อมูลได้เป็นจำนวนมากเช่นกัน การคิดค้นหลักและวิธีการเกี่ยวกับงานทางด้านปัญญาประดิษฐ์เป็นไปอย่างกว้างขวาง จนกล่าวได้ว่าโอกาสของการประยุกต์คอมพิวเตอร์เข้ามาในงานด้านการแปลภาษาจะมีทางเป็นไปได้มากขึ้น

เหตุผลที่สำคัญของผู้พัฒนางานแปลภาษาด้วยเครื่อง [2] ได้แก่

1. ต้องการเอาชนะความหลากหลายของภาษาที่ใช้
2. ราคาและความต้องการของการแปลภาษาสูงขึ้น ๆ
3. งานการประมวลผลภาษารวมชาติเป็นงานที่ทำทลายความสามารถของมนุษย์ และทำให้มนุษย์สามารถสร้างสิ่งใหม่ได้อีกมาก
4. มีความต้องการสร้างเครื่องจักรให้มีความฉลาดและสามารถโต้ตอบกับมนุษย์ได้ดีขึ้น
5. งานที่เป็นผลพลอยได้จากงานแปลภาษามีอีกมาก

1.2 วัตถุประสงค์

พัฒนาระบบไวยากรณ์และพจนานุกรม ภายใต้ระบบทาบเลสตรี (Tapestry) เพื่อให้ระบบสามารถใช้ในการแปลภาษาอังกฤษเป็นภาษาไทยได้

1.3 ขอบเขตของงานวิจัย

- 1.3.1 ระบบที่สร้างขึ้นนี้ใช้กับเครื่องไมโครคอมพิวเตอร์ หน่วยความจำแรมอย่างน้อย 16 เมกะไบต์ ฮาร์ดดิสต์ไม่ต่ำกว่า 20 เมกะไบต์
- 1.3.2 จะใช้ระบบทาบเลสตรีเป็นเครื่องมือในการพัฒนา
- 1.3.3 ฐานข้อมูลพจนานุกรมเป็นเพียงตัวอย่างคำศัพท์จำนวนหนึ่งเท่านั้น
- 1.3.4 จะใช้ทดสอบกับประโยคง่าย ๆ ทีละประโยค

1.4 ขั้นตอนการดำเนินงาน

- 1.4.1 ศึกษาวิธีการทำงานของระบบและโมดูลต่าง ๆ ของระบบทหาเปสตร์
- 1.4.2 ศึกษากฎเกณฑ์ โครงสร้าง และไวยากรณ์ต่าง ๆ ของภาษาอังกฤษและภาษาไทย
- 1.4.3 สร้างระบบไวยากรณ์สำหรับการแปลอังกฤษเป็นไทย
- 1.4.4 สร้างพจนานุกรมของทั้งสองภาษาเพื่อใช้ในระบบทหาเปสตร์
- 1.4.5 ทดสอบโปรแกรม
- 1.4.6 สรุปผลการดำเนินงาน
- 1.4.7 นำเสนอรายงานในรูปวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 ได้เป็นต้นแบบในการพัฒนาระบบการแปลภาษาด้วยคอมพิวเตอร์
- 1.5.2 ได้ระบบไวยากรณ์และพจนานุกรมภาษาอังกฤษและภาษาไทย