

“คอมพิวเตอร์ช่วยแปลภาษา”



*ศิริภัทรา เหมือนมาลัย

๖ รื่องของการแปลภาษาด้วยคอมพิวเตอร์ เริ่มมีมาสามสิบกว่าปีแล้ว แต่ได้หาย

จากไปด้วยความขัดข้องทางเทคนิคทั้งในด้านคอมพิวเตอร์และในด้านของภาษาศาสตร์ ปัจจุบันกำลังกลับมาเป็นที่สนใจกันอีกครั้งหนึ่ง เพราะคอมพิวเตอร์และภาษาศาสตร์ได้เจริญเติบโตคู่กันขึ้นมาถึงจุดหนึ่งทั้งงานในด้านนี้น่าจะเป็นไปได้

จะขอเล่าตั้งแต่ต้นว่า การแปลภาษาโดยใช้คอมพิวเตอร์นั้นคืออะไร มีความเป็นไปได้เพียงใด จากนั้นจะเป็นเรื่องของเทคนิคในระบบต่าง ๆ

การแปลภาษาด้วยเครื่องคอมพิวเตอร์ เรียกอีกอย่างหนึ่งว่า การแปลโดยอัตโนมัติ (Au-

tomatic Translation) ซึ่งหมายถึงการแปลภาษาธรรมชาติจากภาษาหนึ่งเป็นอีกภาษาหนึ่งโดยใช้คอมพิวเตอร์เป็นเครื่องมือ เช่นแปลภาษาอังกฤษเป็นภาษาไทย ภาษาไทยเป็นภาษาญี่ปุ่น เป็นต้น ที่บอกว่า เป็นภาษาธรรมชาติ เพราะปัจจุบันมีภาษาอีกชนิดหนึ่งเรียกว่า Artificial Language เกิดขึ้น เช่น ภาษาโคบอล ภาษาฟอร์แทรน ภาษาเบสิก เป็นต้น การแปลภาษาด้วยคอมพิวเตอร์ที่จะกล่าวถึงในบทความนี้จะไม่เกี่ยวกับภาษาเหล่านี้ เพราะจะนำมาใช้เป็นเครื่องมือในการเขียนโปรแกรมเท่านั้น

เทคโนโลยีการใช้คอมพิวเตอร์แปลภาษา อาจกล่าวได้เป็น 4 เรื่องดังนี้

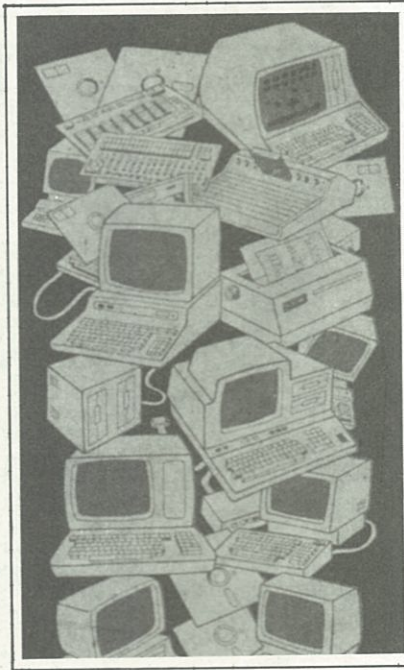
1. ประเภทของระบบระบบการแปลภาษาด้วยคอมพิวเตอร์สามารถแบ่งได้ 3 ประเภทคือ

ก) การแปลภาษาด้วย

เครื่อง (Machine Translation) ระบบนี้จะทำหน้าที่แปลโดยไม่ใช้คนช่วย แต่อาจมีการจัดข้อความที่ต้องการแปลก่อน เรียกว่ามี Pre-editing แต่ไม่ถึงกับช่วยแยกประโยค หรือช่วยแก้ความกำกวมให้เครื่อง เพราะถ้าต้องเสียเวลาทำให้เครื่องมากขนาดนั้น ใช้คนแปลเลยจะดีกว่า นอกจากนั้นอาจมีการตรวจแก้ผลการแปลได้ เรียกว่ามี Post-editing โดยเทียบกับพฤติกรรมมนุษย์ คือ ถ้าใช้คนแปล ก็ต้องมีการตรวจทานแก้ไข ฉะนั้นเมื่อเครื่องแปลแล้วก็ควรจะได้รับ การตรวจทานแก้ไขบ้าง

ข) การแปลภาษาโดยใช้เครื่องช่วย (Machine Aided Translation) วิธีนี้แบ่งได้เป็น

2 ระบบย่อย คือ ระบบการแปลด้วยเครื่องโดยมีคนช่วย กับระบบคนแปลโดยมีเครื่องช่วย ระบบแรกจะใช้เครื่องทำหน้าที่แปล แต่มีคนช่วยเป็นจุด ๆ เช่น ช่วยแก้ความกำกวมของคำ หรือความหมาย หรือให้ออกเครื่องว่าควรนำประโยคไปต่อที่ใด หรือช่วยเลือกคำจากพจนานุกรมให้เลย ส่วนระบบหลัง จะใช้คนทำหน้าที่แปล แต่ให้เครื่องช่วยบางจุด เช่น ให้ช่วยค้นพจนานุกรม จัดรูปแบบของคำ ทั้งสองระบบนี้ไม่มี Pre-editing แต่อาจมี Post-editing ได้



ค) การสร้างคลังศัพท์ (Terminology Data bank) ระบบนี้มักใช้งานก่อนให้คนแปล บางครั้งอาจไม่ใช้คอมพิวเตอร์ช่วยก็ได้ ส่วนมากมักเป็นศัพท์เทคนิค ประโยชน์ของระบบนี้จะอยู่ที่ความทันสมัยนั่นเอง

2. จุดประสงค์ของการใช้ผลการแปล จุดประสงค์ของการแปลมีอยู่ 2 อย่าง คือ การหาสารสนเทศกับการเผยแพร่สารสนเทศ ประการแรก คือ การหาความรู้จากข้อมูลที่มีอยู่จะต้องใช้เครื่องช่วยเพราะอ่านไม่ทัน โดยเฉพาะอย่างยิ่งจากภาษาต่างประเทศ ซึ่งบางทีเนื้อเรื่องอาจไม่น่าสนใจก็ได้ แต่ถ้าไม่เสียเวลาอ่านก็จะไม่ทราบว่าไม่น่าสนใจ ปัจจุบันสารสนเทศเกิดขึ้นอย่าง

มากมาย หากมีวิธีแปลที่รวดเร็วและไม่แพงเกินไป แม้จะมีคุณภาพต่ำก็ยังพอหาความรู้ได้ อย่างน้อยก็จะพอช่วยตัดสินใจได้ว่างานนั้น ๆ ควรนำมาแปลอย่างไรละเอียดหรือไม่ ส่วนในเรื่องของการเผยแพร่สารสนเทศ จะเป็นการเผยแพร่ความรู้จากคนกลุ่มหนึ่งไปสู่คนอีกกลุ่มหนึ่ง เช่น เมื่อประเทศหนึ่งต้องการส่งสินค้าทางด้านเทคโนโลยีไปขายอีกประเทศหนึ่ง ก็จำเป็นต้องมีคู่มือเป็นภาษาของลูกค้าที่จะเข้าใจด้วย การแปลจึงเป็นเรื่องสำคัญในด้านการค้าขายกับต่างประเทศ คู่มือของสินค้าก็จะต้องแปลอย่างระมัดระวัง ต้องถูกต้องและต้องชัดเจน ซึ่งนักแปลเก่ง ๆ ค่าตัวมักจะแพง และหาได้ยาก งานแปล

ก็อาจออกมาช้าไม่ทันใช้งาน

3. จุดประสงค์ของการใช้ระบบการแปล แม้ว่าการแปลงานด้านวรรณกรรม จะถือเป็นการเผยแพร่ก็ตาม แต่มีใช้งานเร่งด่วนไม่เหมือนการแปลงานด้านเทคโนโลยี คนแปลงานด้านวรรณกรรมยังหาได้ง่ายกว่า ประกอบกับคอมพิวเตอร์ก็ไม่น่าหนักในการแปลวรรณกรรม ยกตัวอย่างเช่น ประโยค “LOVE IS BLUE” แต่คอมพิวเตอร์จะแปลประโยคเช่น “OXYGEN IS GAS” ได้ งานแปลด้านเทคโนโลยี นอกจากจะเป็นที่ต้องการกันมากแล้ว ยังต้องมีความสม่ำเสมอในการใช้ศัพท์เทคนิคอีกด้วย นักแปลในด้านนี้ก็หาได้ยาก ดังนั้นระบบการแปลไม่ว่าจะเป็นระบบใด หรือแม้แต่ WORD PROCESSOR ก็นับว่ามีส่วนช่วยได้ทั้งสิ้น การแปลวรรณกรรมจะเน้นที่ลีลาสำนวน แต่การแปลด้านเทคโนโลยีจะเน้นที่เนื้อหา ระบบการแปลภาษาด้วยคอมพิวเตอร์อาจจะขาดลีลาขาดความไพเราะ แต่จะดีในด้านศัพท์และสาระสำคัญของเรื่อง

4. เทคนิคทางด้านภาษาศาสตร์ เทคนิคทางด้านภาษาศาสตร์ที่ใช้กันอยู่ในการแปลภาษาด้วยคอมพิวเตอร์มีอยู่ด้วยกันหลายแบบ อาจจะพอเปรียบเทียบได้เป็นคู่ ๆ ดังนี้

ก) การแปลโดยตรง หรือ การแปลโดยอ้อม

การแปลโดยตรง คือ การแปลจากภาษาหนึ่งไปสู่อีกภาษาหนึ่งโดยเฉพาะ

การแปลโดยอ้อม คือ การแปลที่จะใช้การวิเคราะห์ภาษาต้นฉบับแล้วจึงสร้างภาษาเป้าหมายขึ้น โดยกระบวนการจะแยกจากกัน จะมีการแก้ความกำกวมของภาษาต้นฉบับก่อน โดยจะยังไม่สนใจว่าภาษาเป้าหมายจะเป็นภาษาใด

ตัวอย่างการแปลโดยตรง เช่น “cat” แปลว่า “แมว” เป็นต้น ส่วนตัวอย่างการแปลโดยอ้อม เช่น คำว่า “heat” ในประโยคต้องแก้ความกำกวมเสียก่อน ว่าเป็นคำนาม หรือคำกริยา เป็นต้น

ข) การใช้ภาษากลาง หรือ การใช้การเปลี่ยน (Transfer) เป็นลักษณะของระบบที่พยายามค้นหาความหมายและหาตัวแทนความหมายมาแสดงโดยไม่ขึ้นกับภาษาใด จากนั้นจึงนำตัวแทนความหมายนี้ไปสร้างเป็นภาษาเป้าหมายขึ้นอีกทีหนึ่ง เรื่องของภาษากลางนี้มีแนวคิดมาจากเรื่องของภาษาสากล (Linguistic Universals) ซึ่งนักภาษาศาสตร์และนักปรัชญายังถกเถียงกันอยู่ว่า ตัวแทนของหน่วยความหมาย (Unit of

Meaning) ซึ่งจะต้องเหมือนกันทุกภาษา โดยไม่สนใจว่าแต่ละภาษาจะมีการแสดงออกหรือโครงสร้างไวยากรณ์เป็นอย่างไร ส่วนการใช้การเปลี่ยนเป็นลักษณะของระบบที่ถือว่าตัวแทนแสดงความหมายของหน่วยไวยากรณ์ เช่น ประโยค ของภาษาต้นฉบับ กับภาษาเป้าหมายจะต่างกัน จะต้องมีส่วนเชื่อมต่อนี้ที่เทียบตัวแทนแสดงความหมายที่เป็นลักษณะเฉพาะของภาษาหนึ่ง แล้วเปลี่ยนให้เป็นตัวแทนของอีกภาษาหนึ่ง เช่น ในประโยคภาษาอังกฤษใช้ It เป็นประธาน แล้วตามด้วย Verb to be “It is a pleasure to be here” จะต้องเปลี่ยนเป็นโครงสร้างของภาษาไทย คือ “ยินดีที่ได้มาอยู่ที่นี่” ไม่ใช่แปลตรง ๆ โดยไม่มีการเปลี่ยนว่า “มันเป็นความยินดี มาอยู่ที่นี่”

ระบบการแปลที่ใช้เทคนิคการใช้ภาษากลาง จะประกอบด้วย การวิเคราะห์และการสร้าง ส่วนระบบการแปลที่ใช้เทคนิคการเปลี่ยนจะประกอบด้วย การวิเคราะห์ การเปลี่ยน และการสร้าง

ค) การพิจารณาแบบเฉพาะที่และการพิจารณาแบบครอบคลุม การพิจารณาแบบเฉพาะที่ถือว่าคำเป็นตัวแทนสำคัญในการวิเคราะห์ อาศัยคำที่ปรากฏ

ข้างหน้าหรือข้างหลังเป็นตัวช่วยตัดสินความหมาย ตัวอย่างเช่น heat ถ้ามี the อยู่ข้างหน้าจะถือเป็นคำนาม ถ้าไม่มี the อยู่ข้างหน้าก็เป็นคำกริยา เป็นต้น ส่วนการพิจารณาแบบครอบคลุมจะอาศัย context ช่วยตัดสินความหมาย อาจจะเป็นการวิเคราะห์ประโยคทั้งประโยค หรือย่อหน้าทั้งย่อหน้าก็ได้ เช่น “เมื่อวานเขาไปโรงเรียน” กับ “วันนี้เขาอยู่บ้าน” ดังนั้น “เมื่อวาน” จะเป็น context ที่บอกให้รู้ว่าทำกริยานั้นในอดีต และ “ตอนนี้” เป็น context บอกให้ทราบว่าทำกริยานั้นในปัจจุบัน เป็นต้น

ระบบการแปลภาษาด้วยเครื่องยุคแรก

ระบบการแปลภาษาด้วยเครื่องยุคแรก มุ่งไปที่การแปลเอกสารด้านวิทยาศาสตร์อย่างคร่าว ๆ ผู้ใช้จะต้องรู้เนื้อหาของวิชาการนั้น ๆ อยู่แล้ว ในยุคแรกเชื่อว่าส่วนสำคัญที่สุดของระบบคือพจนานุกรมสองภาษา เช่น อังกฤษ-ฝรั่งเศส ฝรั่งเศส-อังกฤษ การสร้างระบบ จึงเป็นการสร้างพจนานุกรม ส่วนวิธีการปรับปรุงการแปล ก็คือ

- การขยายพจนานุกรมให้รวมทั้งสำนวนและส่วนของประโยคที่มีคำแปลโดยเฉพาะ
- การสร้างโปรแกรมคอมพิวเตอร์ เพื่อจัดความกำกวม

ของคำ การเรียงคำในภาษาเป้าหมาย เป็นต้น

- การเขียนรหัสคำสั่งไว้ในพจนานุกรม เพื่อนำไปเรียกโปรแกรมพิเศษมาทำงานกับคำนั้น ๆ

วิธีนี้ได้เชื่อมความนิยมลง เพราะเมื่อข้อความที่ต้องการแปลยาวมากขึ้น คุณภาพก็เริ่มเลวลงจึงต้องเพิ่มโปรแกรมเพื่อปรับปรุงระบบ จนในที่สุดก็ปรับปรุงต่อไปไม่ไหว

ระบบการแปลภาษาด้วยเครื่องในยุคที่สอง

ในยุคนี้เริ่มมีการตรวจหาลักษณะโครงสร้างของภาษา แล้วสร้างรูปแบบต่าง ๆ ขึ้น เพราะนักภาษาศาสตร์เริ่มพิจารณาภาษาธรรมชาติ, นักคณิตศาสตร์เริ่มพัฒนาทฤษฎีด้านภาษารูปนัย, นักคอมพิวเตอร์เริ่มรู้จักแนวคิดขั้นพื้นฐานของโปรแกรม และเริ่มหาทฤษฎีการแปลโปรแกรม ในยุคนี้เองได้มีการแยกซอฟต์แวร์ออกเป็น 2 ส่วน คือ ซอฟต์แวร์ที่เป็นโปรแกรมคอมพิวเตอร์ กับซอฟต์แวร์ทางด้านภาษาศาสตร์ โครงสร้างของประโยคกับพจนานุกรมจะแยกออกจากกัน ซอฟต์แวร์ด้านภาษาศาสตร์จะถือไวยากรณ์และพจนานุกรมซอฟต์แวร์ในส่วนที่เป็นโปรแกรมคอมพิวเตอร์ที่จะใช้ทำงานกับซอฟต์แวร์ด้านภาษาศาสตร์จะ

ไม่ขึ้นอยู่กับภาษาใดภาษาหนึ่ง โดยเฉพาะ

ระบบการแปลภาษาด้วยเครื่องตามแนวคิดนี้ เกิดขึ้นมากมาย นักวิจัยต่างก็พยายามวิเคราะห์ว่า ขั้นตอนการแปลนั้นควรเป็นอย่างไร แล้วสร้างซอฟต์แวร์คอมพิวเตอร์ขึ้น ให้พร้อมจะทำงานกับซอฟต์แวร์ด้านภาษาศาสตร์ นักคอมพิวเตอร์จากหลายกลุ่มวิจัยได้สร้างซอฟต์แวร์คอมพิวเตอร์ดังกล่าวเรียบร้อยแล้วและกำลังปรับปรุงให้ใช้ได้สะดวกยิ่งขึ้น ซอฟต์แวร์คอมพิวเตอร์จะแบ่งได้เป็น 2 พวก คือ พวกที่ใช้ระบบการเปลี่ยนกับพวกที่ใช้ภาษากลาง ซึ่งในยุคแรกที่ยังไม่มีการวิเคราะห์ภาษาต้นฉบับ การแปลจะมีลักษณะเป็นการเปลี่ยนโดยตรงจากภาษาหนึ่งเป็นอีกภาษาหนึ่ง ระบบหนึ่งจะทำได้เพียงคู่เดียว ต่อมาเมื่อมีการวิเคราะห์ภาษาต้นฉบับ

และการสร้างภาษาเป้าหมาย ความแตกต่างของแต่ละระบบจะอยู่ที่ความลึกซึ้งของการวิเคราะห์ ช่วงนี้การแปลจะอยู่ในลักษณะหลายภาษา คือ จะแปลภาษาคู่ใดก็ได้ เพราะการวิเคราะห์กับการสร้างไม่ขึ้นต่อกัน ระบบที่ใช้การเปลี่ยนจะเริ่มเปลี่ยนที่จุดใดก็ได้แล้วแต่ความลึกของการวิเคราะห์ถ้าวิเคราะห์ได้ลึกมาก การเปลี่ยนก็ซับซ้อนน้อยลง เมื่อใดที่วิเคราะห์ได้ลึกถึงขนาดที่สามารถเข้าถึงภาษากลางได้ เมื่อนั้นก็จะไม่มีการเปลี่ยน เพราะจะสร้างประโยคเป้าหมายจากภาษากลางได้ทันที นี่จึงเป็นระบบที่ใช้ภาษากลาง

อย่างไรก็ตาม แนวคิดในเรื่องความลึกของการวิเคราะห์อย่างไหนจึงเรียกว่าลึก และภาษากลางที่ค้นพบนี้คือ ภาษากลางจริงหรือ จึงเป็นสิ่งที่น่าติดตามกันต่อไป □ ■

