

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การแปลภาษาด้วยเครื่อง (Machine Translation) คือ การนำระบบอัตโนมัติที่ทำงานบนเครื่องคอมพิวเตอร์มาใช้ในการแปลข้อความจากภาษา หนึ่ง ไปเป็นอีกภาษาหนึ่ง โดยเมื่อป้อนข้อความภาษาต้นฉบับเข้าไปในระบบ โปรแกรมจะทำการวิเคราะห์ข้อมูลภาษาต้นฉบับ เลือกรูปแบบ และสร้างข้อความของภาษาเป้าหมายออกมา ช่วยให้สามารถแปลข้อความได้เป็นจำนวนมาก และรวดเร็ว หัวข้อวิจัยเรื่องการแปลภาษาด้วยเครื่องจัดเป็นงานแขนงหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing) งานวิจัยด้านนี้ถูกเผยแพร่สู่สาธารณะครั้งแรกในปี พ.ศ. 2497 โดยมหาวิทยาลัยจอร์จทาวน์ (Georgetown University) ที่ระบบสามารถแปลข้อความจากภาษารัสเซียเป็นภาษาอังกฤษได้จำนวน 49 ประโยค มีจำนวนคำศัพท์ในพจนานุกรม 250 คำและกฎไวยากรณ์ 6 กฎ [1]

ระบบการแปลภาษาด้วยเครื่องในปัจจุบันได้ถูกนำไปใช้กับการแปลภาษาต่างๆ จำนวนมาก ไม่ว่าจะเป็นภาษาที่มีกฎไวยากรณ์ใกล้เคียงกัน หรือแตกต่างกันมาก แต่ในประเทศไทย การแปลภาษาด้วยเครื่องของภาษาไทยยังมีอยู่น้อยมากเมื่อเทียบกับต่างประเทศ เช่น อเมริกา จีน ญี่ปุ่น ออสเตรเลีย และประเทศในแถบยุโรป และส่วนมากจะเป็นการแปลระหว่างภาษาไทยและภาษาอังกฤษ เช่น ระบบแปลภาษาไทย-ภาษาอังกฤษออนไลน์ ที่ชื่อว่า “ภาษิต” (PARSIT) โดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ หรือระบบแปลภาษาด้วยเครื่องจากภาษาอังกฤษเป็นภาษาไทย ที่ชื่อว่า “แปลไทย” เป็นต้น

ประเทศญี่ปุ่นซึ่งเป็นประเทศหนึ่งในโลกที่มีความก้าวหน้าเป็นอย่างมากทั้งทางด้านอุตสาหกรรมและด้านเทคโนโลยี มีบริษัทสัญชาติญี่ปุ่นจำนวนมากในประเทศต่างๆ ทั่วโลก มีนักท่องเที่ยวชาวญี่ปุ่นจำนวนมากเดินทางไปท่องเที่ยวรอบโลก และมีผู้คนจากประเทศต่างๆ ทั่วโลกให้ความสนใจในการเรียนภาษาญี่ปุ่น จากเหตุผลต่างๆ เหล่านี้ทำให้ภาษาญี่ปุ่นซึ่งเป็นภาษาราชการที่คนญี่ปุ่นใช้ตั้งแต่ในอดีตจนถึงปัจจุบัน ได้รับความสนใจจากนักวิจัยจำนวนหนึ่ง มีการพัฒนาระบบการแปลภาษาของประเทศตนเองไปเป็นภาษาญี่ปุ่น เช่น การแปลภาษาอังกฤษเป็นภาษาญี่ปุ่น [2-5] การแปลภาษาจีนเป็นภาษาญี่ปุ่น [6-7] หรือการแปลภาษาเกาหลีเป็นภาษาญี่ปุ่น [8-9]

ถึงแม้จะมีนักเรียนไทยจำนวนมากสนใจศึกษาภาษาญี่ปุ่น และมีคนญี่ปุ่นมาท่องเที่ยวหรือทำงานในประเทศไทยเป็นจำนวนมากก็ตาม แต่ก็ยังไม่มีโปรแกรมประยุกต์หรืองานวิจัยชิ้นใดที่เกี่ยวข้องกับการแปลภาษาไทยเป็นภาษาญี่ปุ่นโดยตรง จะมีก็เพียงแต่บริการแปลภาษาบนอินเทอร์เน็ตของ Google Translate (<http://translate.google.com>) ที่สามารถแปลภาษาไทยเป็นภาษาญี่ปุ่นได้ แต่ยังมีคามถูกต้องน้อย และไม่สามารถแปลประโยคที่มีความซับซ้อนได้

1.2 แนวทางการแก้ปัญหา

งานวิจัยนี้มีวัตถุประสงค์หลัก คือการพัฒนาาระบบแปลภาษาไทยเป็นภาษาญี่ปุ่นโดยตรง โดยเริ่มต้นจากการพัฒนาระบบแปลภาษาด้วยเครื่องแบบใช้ ฐานกฎ (Rule-based Machine Translation) ซึ่งเป็นหลักการที่มักจะนำมาใช้ในการแปลและวิเคราะห์ความยากง่ายในการแปลคู่ภาษานั้นๆ

ในมุมมองด้านภาษาศาสตร์ ภาษาไทยและภาษาญี่ปุ่นมีความแตกต่างกันมากเชิงโครงสร้าง เช่น ภาษาญี่ปุ่นมีการจัดเรียงลำดับของคำแตกต่างจากภาษาไทย ภาษาญี่ปุ่นมีคำช่วยที่ใช้ระบุประเภทและหน้าที่ของคำ คำกริยาในภาษาญี่ปุ่นแต่ละคำจะใช้คำช่วยแตกต่างกัน และมีกฎการผันกริยาที่ขึ้นอยู่กับคำช่วยกริยาจำนวนมาก เป็นต้น ทำให้การแปลนั้นจำเป็นจะต้องรู้กฎไวยากรณ์ และคู่คำศัพท์ของทั้งสองภาษาเป็นอย่างดี

ในงานวิจัยนี้จะนำเสนอหลักการแปลภาษาไทยเป็นภาษาญี่ปุ่น โดยจะเน้นเรื่องการใช้คำช่วยกริยานุเคราะห์ และการผันกริยา ให้ถูกต้องเป็นหลัก และจะทำการพัฒนาฐานข้อมูลคู่คำศัพท์ของตนเอง โดยนำคำศัพท์ภาษาไทยมาจากคลังคำศัพท์ Orchid Corpus [10] ส่วนประกอบของฐานข้อมูลคู่คำศัพท์นี้ ประกอบด้วยคำศัพท์ของทั้งสองภาษา และหน้าที่ของคำศัพท์ในประโยค (Part of Sentence: POS)

1.3 สรุปสาระสำคัญจากเอกสารที่เกี่ยวข้อง

ในปัจจุบันเมื่อการสื่อสารไม่ได้จำกัดอยู่แค่ภายในประเทศตัวเอง มีการติดต่อกันในทุกๆ ด้านกับต่างประเทศ ไม่ว่าจะเป็นด้านเศรษฐกิจ เทคโนโลยี และการเมือง ทำให้การได้รับข้อมูลที่ต้องการเป็นสิ่งที่มีความจำเป็น ดังนั้นการแปลภาษาจึงเป็นหัวข้อที่นักวิจัยจำนวนมากให้ความสนใจ และมีนักวิจัยส่วนหนึ่งให้ความสนใจกับการแปลภาษาของประเทศตนเองไปเป็นภาษาญี่ปุ่น ซึ่งถือว่าเป็นประเทศที่มีอิทธิพลมากประเทศหนึ่งของโลก

Yin และคณะ [6] ได้นำเสนอหลักการแปลภาษาด้วยเครื่องแบบตามลำดับขั้นแนวใหม่ สำหรับการวิเคราะห์ประโยคภาษาจีนที่ยาวและซับซ้อน สำหรับระบบแปลภาษาด้วยเครื่องจากภาษาจีนเป็นภาษาญี่ปุ่น โดยวิเคราะห์จากการเว้นวรรค คำเชื่อมประโยค และไวยากรณ์ภาษาจีน

วิธีการวิเคราะห์ไวยากรณ์วิเคราะห์จากลักษณะของไวยากรณ์ของตัวอย่างประโยคภาษาจีนที่รวมลักษณะทางไวยากรณ์ ความยาวของประโยค การเว้นวรรค และคำประเภทต่างๆ ขั้นตอนแรกจะเป็นการวิเคราะห์คำเชื่อมประโยค และทำการแบ่งประโยคที่ยาวและซับซ้อนออกเป็นส่วนย่อยๆ จากนั้นจึงเริ่มแปลแต่ละส่วนย่อย แล้วนำมารวมกันเป็นประโยคผลลัพธ์โดยหลักการจัดเรียงลำดับคำแบบครอบคลุมทั้งประโยค ซึ่งวิธีการนี้สามารถแก้ปัญหาเรื่องเวลาที่ใช้ในการแปลและปัญหาความกำกวมของภาษาจีนได้ และมีความถูกต้องในการแปล 89.4% จากการทดลองกับประโยคภาษาจีนที่ยาวและซับซ้อน 1000 ประโยค

ส่วนในประเทศไทย งานวิจัยเกี่ยวกับการพัฒนาระบบแปลภาษาด้วยเครื่อง จะมุ่งเน้นไปที่การแปลระหว่างภาษาไทยและภาษาอังกฤษ

ธีรพงษ์และคณะ [11] ได้พัฒนาระบบแปลภาษาไทย-ภาษาอังกฤษออนไลน์ ที่ชื่อว่า “ภายิต” (PARSIT) ซึ่งใช้หลักการแปลแบบกฎไวยากรณ์และภาษากลาง โดยมุ่งเน้นที่กระบวนการก่อนทำการแปลและขั้นตอนการวิเคราะห์กฎไวยากรณ์ ซึ่งเป็นส่วนที่สำคัญของปัญหาด้านไวยากรณ์ในภาษาไทย งานวิจัยนี้ได้เผยแพร่ให้ใช้งานได้ผ่านทางเว็บไซต์โดยไม่เก็บค่าใช้จ่าย แต่ในปัจจุบันเว็บไซต์ดังกล่าวไม่สามารถใช้งานได้แล้ว

งานวิจัยที่สามารถแปลภาษาไทยเป็นภาษาอื่นๆ นอกเหนือจากภาษาอังกฤษได้ เช่น งานวิจัยเกี่ยวกับการแปลภาษาด้วยเครื่องแบบหลายภาษา (Multi-Lingual Machine Translation) [12] ซึ่งงานวิจัยนี้เป็นโครงการความร่วมมือระหว่างรัฐบาลไทยและรัฐบาลญี่ปุ่น วัตถุประสงค์ของโครงการนี้คือการพัฒนาและพัฒนาระบบแปลภาษาที่สามารถแปลเป็นภาษาอื่น ๆ ได้ โดยใช้ภาษากลาง ซึ่งเป็นภาษาที่ถูกออกแบบมาเพื่อใช้เป็นสื่อกลางในการแปล โดยที่นักวิจัยฝ่ายไทยจะทำการออกแบบและพัฒนาระบบการแปลภาษาไทยไปเป็นภาษากลาง และนำภาษากลางมาสร้างเป็นภาษาไทยอีกทอดหนึ่ง สำหรับนักวิจัยฝ่ายญี่ปุ่นก็จะทำการวิจัยเช่นเดียวกันนี้ในภาษาญี่ปุ่น และเมื่อนำระบบทั้งสองมารวมเข้าด้วยกัน จะสามารถแปลภาษาไทยเป็นภาษาญี่ปุ่น หรือจากภาษาญี่ปุ่นเป็นภาษาไทยได้ แต่โครงการนี้ประสบผลสำเร็จเพียงแค่ระดับหนึ่งเท่านั้น ระบบที่พัฒนาขึ้นไม่ได้นำออกมาใช้จริง และไม่มีรายงานการวิเคราะห์เรื่องความถูกต้องในการแปลภาษาไทยเป็นภาษาญี่ปุ่นอย่างชัดเจน

เทพชัยและคณะ [13] ได้นำเสนองานวิจัยที่มีแนวคิดแบบเดียวกันกับงานวิจัยข้างต้น แต่ใช้ภาษาอังกฤษเป็นภาษากลาง โดยจะทำการแปลภาษาไทยเป็นภาษาอังกฤษ แล้วแปลภาษาอังกฤษเป็นภาษาญี่ปุ่นอีกทอดหนึ่ง แต่งานวิจัยนี้ก็ยังเป็นเพียงแค่การนำเสนอแนวคิด และการทดลองพื้นฐานเพื่อพัฒนาระบบที่สามารถนำเครื่องมือที่มีอยู่ในปัจจุบัน มาประยุกต์ใช้ในการแปลภาษาไทยเป็นภาษาอื่นๆ ที่นอกเหนือจากภาษาอังกฤษได้

1.4 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและ พัฒนาระบบ แปลภาษา ไทยเป็นภาษา ญี่ปุ่น ซึ่งใช้วิธีการแปลภาษาด้วย เครื่องแบบใช้กฎไวยากรณ์ และเพื่อให้เป็นต้นแบบและแนวทางในการพัฒนาระบบแปลภาษา ระหว่างภาษาไทยและภาษาญี่ปุ่นต่อไปในอนาคต

1.5 ขอบเขตการทำวิจัย

สามารถแปลข้อความภาษาไทยที่เป็นคำศัพท์ วลี ประโยคความเดียว และประโยคความ รวม ที่เป็นประโยคพื้นฐานในบทสนทนาภาษาไทย ซึ่งข้อความที่สามารถแปลได้นั้น ต้องผ่าน กระบวนการตัดคำที่ทำการตัดคำได้อย่างถูกต้องโดยประยุกต์ใช้จากโปรแกรม Swath [14] จากนั้น ทำการวิเคราะห์คำศัพท์และประโยค ด้วยกฎไวยากรณ์ที่เขียนด้วยภาษาไบสัน (Bison) [15] และ เมื่อแปลออกมาแล้ว จะได้ประโยคภาษาญี่ปุ่นที่มีความยากของไวยากรณ์และคำศัพท์อยู่ใน ระดับพื้นฐาน ระดับ 3 ตามมาตรฐาน การสอบวัดระดับความสามารถทางภาษาญี่ปุ่น (Japanese Language Proficiency Test)

1.6 ประโยชน์ที่จะได้รับจากการวิจัย

ได้โปรแกรมคอมพิวเตอร์ที่สามารถแปลภาษา ไทยเป็นภาษาญี่ปุ่น ซึ่งนักเรียน นักศึกษา และผู้สนใจภาษาญี่ปุ่นสามารถนำไปใช้เพื่อช่วยในการศึกษาภาษาญี่ปุ่นได้

1.7 ระเบียบวิธีการทำวิจัย

1.7.1 ศึกษางานวิจัย รวบรวมข้อมูลที่เกี่ยวข้องกับการแปลภาษาด้วยเครื่อง เพื่อให้ทราบถึงการทำงาน และแนวทางในการพัฒนาระบบแปลภาษาด้วยเครื่อง

1.7.2 ศึกษาทฤษฎี และขั้นตอนวิธีพื้นฐานที่ใช้ในงานแปลภาษาด้วยเครื่อง รวมถึงกฎไวยากรณ์ภาษาไทยและภาษาญี่ปุ่น

1.7.3 จัดเตรียมและเก็บข้อมูลที่จะนำมาใช้สำหรับการวิจัยและการทดสอบระบบ โดยการสอบถามจากอาจารย์ผู้เชี่ยวชาญภาษาญี่ปุ่น และจากหนังสือด้านภาษาศาสตร์ที่เกี่ยวข้อง

1.7.4 ออกแบบ และพัฒนาระบบการแปลภาษาด้วยเครื่อง โดยวิเคราะห์ขั้นตอนวิธีการสร้างกฎไวยากรณ์และต้นไม้เชิงโครงสร้างของประโยคภาษาไทยและภาษาญี่ปุ่น รวมไปถึงการจัดเรียงประโยคให้ถูกต้อง และการกำหนดข้อยกเว้นต่างๆ

1.7.5 ทดสอบการทำงานของระบบ โดยวัดประสิทธิภาพของการแปลด้วยวิธีการที่เป็นมาตรฐาน และวัดความถูกต้องและความพึงพอใจในการแปล โดยการทำแบบสอบถามกับอาจารย์ผู้เชี่ยวชาญภาษาญี่ปุ่น

1.7.6 วิเคราะห์ วิจารณ์ และสรุปผลการวิจัย

1.7.7 จัดทำและเสนอรายงานวิทยานิพนธ์



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่

Copyright© by Chiang Mai University

All rights reserved