



ความสามารถในการแปลความหมายระดับข้อความของ Google Translate ระบบ Neural Machine Translation จากภาษาญี่ปุ่นเป็นภาษาไทย¹

ณัฐริชา ทับทิม *

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม ประเทศไทย

The Capacity of Contextual Meaning Translation of Google Translate in the Form of Neural Machine Translation System from Japanese to Thai

Natthira Tuptim *

Faculty of Humanities and Social Sciences, Pibulsongkram Rajabhat University, Thailand

Article Info

Research Article

Article History:

Received 22 April 2020

Revised 29 May 2020

Accepted 3 June 2020

คำสำคัญ

การแปล

ภาษาระดับข้อความ

Google Translate

ระบบ Neural Machine

Translation

Keywords:

Translation

Contextual Meaning

Google Translate in Neural

Machine Translation System

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความสามารถในการแปลความหมายระดับข้อความของแอปพลิเคชัน Google Translate ระบบ Neural Machine Translation รายการภาษาความหมายระดับข้อความประกอบไปด้วยคำประสม คำนามและคำประสมคำกริยา จำนวน 136 คำ คำปรากฏร่วมประเภท คำนาม+คำกริยา จำนวน 100 คำ จำนวนที่มีคำนามอวัยวะเป็นส่วนประกอบ จำนวน 100 จำนวน และประโยคที่ประกอบจากรูปไวยากรณ์ระดับ N4 จำนวน 100 ประโยค รวมทั้งหมด 436 รายการ โดยแปลจากภาษาญี่ปุ่นเป็นภาษาไทย สำนวนจะระหว่างเดือนตุลาคมถึงพฤศจิกายน 2561 และตรวจสอบความสามารถการแปลโดยผู้เชี่ยวชาญภาษาญี่ปุ่นและภาษาไทย จำนวน 3 คน ผลการวิจัยพบว่า Google Translate ระบบ Neural Machine Translation แปลความหมายจากภาษาญี่ปุ่นเป็นภาษาไทยอยู่ในระดับไม่เป็นที่ยอมรับคิดเป็นร้อยละ 73.39 และแปลความหมายจากภาษาญี่ปุ่นเป็นภาษาไทยอยู่ในระดับเป็นที่ยอมรับคิดเป็นร้อยละ 26.61 โดยรายการภาษาที่ Google Translate ระบบ Neural Machine Translation สามารถแปลได้ในระดับยอมรับได้มากที่สุดเรียงจากมากไปหาน้อยได้แก่ คำปรากฏร่วม คำประสมคำนาม คำประสมคำกริยา ประโยค และสำนวน ตามลำดับ

* Corresponding author

E-mail address:

ntuptim@hotmail.com

Abstract

This research aims to study the capacity of contextual meaning translation of Google Translate in the form of Neural Machine Translation system. A Japanese language package in a contextual level meaning comprises 436 items which can be divided into 136 words of compound nouns and compound verbs, 100 noun + verb collocations, 100 idioms about organs, and 100 sentences consisting of N4 level grammatical forms. Data collection proceeded from October to November 2018, and the capacity of translation was examined by three experts of Japanese and Thai. The results revealed that the translation capacity from Japanese to Thai of Google Translate in the form of Neural Machine Translation system was at an unacceptable level at 73.39 percent while the acceptability was at 26.61 percent. The acceptability of translation capacity of Google Translate in the form of Neural Machine Translation system for the language package arranging in descending order was collocations, compound nouns, compound verbs, sentences, and idioms, respectively.

1. บทนำ

ความเจริญก้าวหน้าทางเทคโนโลยีทำให้ปัจจุบันคอมพิวเตอร์และอุปกรณ์อิเล็กทรอนิกส์เข้ามามีบทบาทในทุกวงการ ความฉลาดและความสามารถเข้าใจในสิ่งที่มนุษย์ต้องการของคอมพิวเตอร์เป็นผลมาจากการพัฒนาของวิทยาการด้านปัญญาประดิษฐ์ (Artificial Intelligence) การประมวลผลภาษามนุษย์ด้วยคอมพิวเตอร์หรือการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นส่วนหนึ่งของงานวิจัยทางด้านปัญญาประดิษฐ์ ที่พยายามจำลองความฉลาดของมนุษย์ให้กับคอมพิวเตอร์อื่นได้แก่ การแก้ปัญหาโดยใช้เหตุผล การเรียนรู้วิธีการแก้ปัญหาด้วยตนเอง รวมทั้งการสื่อสารด้วยภาษามนุษย์ การประยุกต์ใช้งานจากการประมวลผลภาษาธรรมชาติที่โดดเด่นและส่งผลกระทบต่อมนุษย์เป็นอย่างมากในปัจจุบัน คือ การแปลภาษาอัตโนมัติ (Machine Translation) (อัศนีย์ ก่อตระกูล และคณะ, 2541, น. 94)

การแปลภาษาอัตโนมัติ (Machine Translation) ในภาษาญี่ปุ่นเรียกว่า Kikaihonyaku (機械翻訳) หรือ Jidōhonyaku (自動翻訳) ในภาษาไทยมีคำเรียก Machine Translation ในรูปแบบอื่นเช่นกัน เช่นเรียกว่า เครื่องจักรช่วยแปล ในงานวิจัยนี้ใช้คำว่า การแปลภาษาอัตโนมัติเมื่อกล่าวถึงลักษณะการแปล และใช้คำว่าเครื่องจักรช่วยแปลเมื่อกล่าวถึงตัวเครื่องมือที่ใช้ในการแปล โดยการแปลภาษาอัตโนมัติในพจนานุกรม Kōjien (2008, p. 659) หมายถึงการแปลภาษาเขียนหรือภาษาสนทนาโดยอัตโนมัติผ่านระบบคอมพิวเตอร์

การพัฒนาทวิวิธีการแปลภาษาอัตโนมัตินี้พัฒนามาอย่างยาวนานจนในปีค.ศ. 2016 ภูเกิลได้ประกาศการนำทวิวิธี Neural Machine Translation (NMT) มาใช้ในการแปลภาษาทำให้การแปลภาษาด้วยระบบปฏิบัติการคอมพิวเตอร์ของภูเกิลคือ Google Translate สามารถแปลภาษาได้ดีมากขึ้นจนมีการคาดการณ์กันว่าน่าจะแปลภาษาจะเป็นหนึ่งในอาชีพที่จะถูกแทนที่ด้วยปัญญาประดิษฐ์เพราะการพัฒนาแบบก้าวกระโดดของ Google Translate (9 อาชีพที่ตกงานก่อนใครเพราะหุ่นยนต์และ AI เข้ามาแทนที่, 2017, 20 มิถุนายน)

จากการพัฒนาเทคโนโลยีข้างต้นทำให้การเรียนการสอนภาษาต่างประเทศจำเป็นต้องมีการปรับตัวเพื่อให้สิ่งที่ถูกคาดการณ์ไม่เกิดขึ้นหรือให้เกิดผลกระทบต่อการประกอบอาชีพน้อยที่สุดและเพื่อให้มนุษย์สามารถใช้ภาษาร่วมกับปัญญาประดิษฐ์ได้ การแปลภาษาโดยมนุษย์จึงจำเป็นต้องเป็นสิ่งที่ปัญญาประดิษฐ์ไม่สามารถทำได้ดีเทียบเท่ากับมนุษย์ กระบวนการสอนการแปลภาษาต่างประเทศในปัจจุบันจึงต้องมีการศึกษาสิ่งที่ปัญญาประดิษฐ์หรือในที่นี้คือ Google Translate ไม่สามารถแปลได้ดีและนำข้อมูลที่ได้มาประยุกต์ใช้เป็นรูปแบบและเทคนิคการสอนการแปลภาษาที่ต้องอาศัยความสามารถของมนุษย์

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความสามารถในการแปลภาษาความหมายระดับข้อความของ Google Translate ระบบ Neural Machine Translation จากภาษาญี่ปุ่นเป็นภาษาไทย

กำหนดขอบเขตการศึกษาคือคำ กลุ่มคำ สำนวนและประโยคที่มีความหมายระดับข้อความ ความหมายระดับข้อความ หมายถึง ความหมายของภาษาเมื่อถูกนำไปใช้จริงในเวลา สถานการณ์ และความสัมพันธ์กับบุคคลรวมถึงความหมายของภาษาที่มีหรือที่เกิดจากกระบวนการทางภาษา ที่ทำให้ความหมายของคำมีความซับซ้อนมากขึ้นไม่สามารถเข้าใจความหมายตามรูปคำได้ โดยผลวิจัยที่ได้จะนำไปประยุกต์ใช้ในการสอนการแปลบูรณาการกับรายวิชาการแปลภาษาญี่ปุ่น และการแปลภาษาญี่ปุ่นแบบสาม

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

ศาสตร์การแปลเป็นกระบวนการที่เกี่ยวกับภาษา เป็นกระบวนการที่ต้องอาศัยเกณฑ์พิเศษนอกภาษาทั้งชุดอีกด้วย เกณฑ์พิเศษนอกภาษาที่ว่่านี้ยกตัวอย่างเช่น วัฒนธรรม ประวัติศาสตร์ วิธีการนับวันเดือนปี ฯลฯ การแปลเฉพาะตัวภาษาโดยไม่มีวัฒนธรรมเกี่ยวข้องด้วยนั้นเป็นเรื่องที่อาจเกิดขึ้นได้ยาก จำต้องประกอบไปด้วย ความหมายของคำ บริบท การใช้ภาษา และวัฒนธรรม (Taylor, 1998, อ้างถึงในสุพรรณิ ปิ่นมณี, 2555, น. 9) อีกทั้งหลักสำคัญของการแปลอีกสิ่งหนึ่งคือความเท่าเทียม (Equivalence) ของต้นฉบับและบทแปล หนึ่งในความเท่าเทียมกันที่ถูกกล่าวถึงคือ Dynamic equivalence คือการเลือกใช้สำนวนให้เหมาะสมกับความคาดหวังทางวัฒนธรรมปลายทาง (สมเกียรติ เชวงกิจวินิช, 2549, น. 22-23)

จากกระบวนการแปลข้างต้นที่จำเป็นต้องประกอบไปด้วยความรู้ภาษา (ที่มีความหมายระดับคำและความหมายระดับข้อความ ความหมายระดับข้อความในที่นี้คือความหมายที่ต้องอาศัยบริบท วัฒนธรรม การใช้ภาษาประกอบความเข้าใจความหมาย) และรวมถึงความรู้เรื่องวัฒนธรรมของภาษาปลายทางทำให้ศาสตร์การแปลเป็นศาสตร์ที่มีความซับซ้อนและต้องอาศัยประสบการณ์ในการแปลให้ถึงหลักการแปลที่เหมาะสมกับวัฒนธรรมของภาษาปลายทาง หรือแบบ Dynamic equivalence

ในปี ค.ศ. 2016 กูเกิลพัฒนา ระบบการแปล Google Neural Machine Translation (GNMT) ระบบ GNMT เป็นระบบที่ให้เครื่องจักรเรียนรู้ พัฒนาตนเองจากการพบคำซ้ำๆ เป็นการสร้างประสบการณ์เพื่อสร้างความรู้ใหม่โดยที่โปรแกรมเมอร์ไม่จำเป็นต้องเขียนโปรแกรมให้ครอบคลุมทุกขั้นตอนการทำงาน ระบบ GNMT สามารถสร้างภาษากลางขึ้นมาได้โดยปัญญาประดิษฐ์จะแปลงจากอีกภาษาหนึ่งมาเป็นภาษากลาง พร้อมกับแปลงไปเป็นอีกภาษาหนึ่งที่ผู้ใช้งานต้องการ ปัจจุบัน GNMT นี้ถูกนำมาใช้ในระบบ Google Translate อย่างเต็มตัวและรองรับการแปลข้ามภาษาแบบอิสระได้ 10 คู่ภาษาจากทั้งหมด 16 คู่ภาษาซึ่งรวมถึงภาษาไทยและภาษาญี่ปุ่น (Techsource Team, 2560, 14 เมษายน)

การพัฒนาเครื่องจักรช่วยแปลมีการพัฒนามาอย่างต่อเนื่อง ในงานของ Gotō (2018, pp. 14-25) อธิบายเทคโนโลยีเครื่องจักรช่วยแปลว่ามีการแบ่งออกได้เป็น 2 ส่วนใหญ่ คือ รูปแบบ Rule-based Machine Translation (RBMT) ซึ่งอาศัยการแปลผ่านกฎของภาษาหรือกฎไวยากรณ์ ที่ถูกใส่ในระบบประมวลผลของคอมพิวเตอร์ อีกรูปแบบหนึ่งคือ รูปแบบ Corpus-based Machine Translation ซึ่งอาศัยการรวบรวมข้อมูลการแปลเป็นภาษาต่างๆ จัดเก็บเป็นคลังข้อมูลภาษาขนาดใหญ่ โดยรูปแบบ Corpus-based Machine Translation ถูกแบ่งเป็นรูปแบบย่อยๆ ได้ เป็น Example-based Machine Translation: EBMT (Yōrē be-su kikaihonyaku: 用例ベース機械翻訳) Statistic Machine Translation: SMT (Tōkē teki kikaihonyaku: 統計的機会翻訳) และ Neural Machine Translation: NMT (Nyu-raru kikaihonyaku: ニューラル機械翻訳)

ระบบ NMT ซึ่งเป็นที่นิยมในการวิจัยพัฒนาและนำมาใช้ช่วงปี 2005 คือการใช้ระบบ Distributed representation, word embeddings (Bunsan hyōgen: 分散表現) การใช้คำศัพท์ที่มีความหมายใกล้เคียงกันให้อยู่ในเวกเตอร์ใกล้เคียงกัน เช่นคำว่า Kodomo (子供) กับ Shōnen (少年) (เวกเตอร์ หรือ Vector คือ เส้นสมมุติหรือลูกศรสมมุติแสดงปริมาณและทิศทางของข้อมูล) และใช้ระบบ RNN หรือ Recurrent Neural Network) ซึ่งเป็นระบบเก็บรักษาประโยคหรือข้อความที่ถูกใส่เข้าไปก่อนหน้านี้ไม่ให้หายไปเป็นลักษณะของ Long Short-Term Memory ปัญหาพื้นฐานของระบบ NMT คือการแปลประโยคขนาดยาวและคำ หรือประโยคที่มีข้อมูลการใช้ความถี่น้อย ก็มีผลทำให้การแปลคำหรือประโยคเหล่านี้ของระบบ NMT ไม่มีประสิทธิภาพไปด้วย นอกจากนี้ระบบ NMT ยังไม่มีการแบ่งส่วนที่ถูกแปลกับส่วนที่ไม่ถูกแปลจึงทำให้เกิดการแปลข้ามหรือแปลไม่ครบที่เรียกว่า Yakunuke (訳抜け) แต่เนื่องจากระบบ NMT เป็นระบบที่ใช้คลังข้อมูลขนาดใหญ่ ทำให้คำแปลของเครื่องจักรช่วยแปลที่ใช้ระบบนี้มีแนวโน้มแปลภาษาที่เป็นธรรมชาติและแปลข้อความขนาดสั้น (เช่นประมาณ 20 คำ) ได้ละเอียดมากขึ้นทั้งการเรียงคำ ซึ่งสะท้อนลักษณะเด่นของระบบ NMT ที่ใช้บริบทแวดล้อมทั้งหมดของข้อมูลที่อยู่ในคลังข้อมูลมาเป็นข้อมูลประกอบการแปล และใช้ Distributed representation, word embeddings ในการเลือกคำศัพท์ที่มีความหมายคล้ายกันจากสถิติทำให้แนวโน้มคำที่ถูกนำมาใช้ในการแปลมีความเหมาะสมมากขึ้น

Bentivogli et al. (2016) ศึกษาเปรียบเทียบความสามารถในการแปลภาษาอัตโนมัติ (Machine Translation) ระหว่างระบบ Phrase-based Machine Translation (PBMT) กับระบบ Neural Machine Translation (NMT) โดยนำระบบทั้งสองมาเป็นเครื่องมือในการแปลภาษาพูดใน TED Talks จากภาษาอังกฤษเป็นภาษาเยอรมันผลปรากฏว่าในแง่คุณภาพการแปล ระบบ NMT แปลได้ดีกว่าระบบ PBMT อย่างมีนัยสำคัญทางสถิติ นอกจากนี้ผลการเปรียบเทียบข้อคิดในการแปลแต่ละด้านระหว่าง 2 ระบบก็ให้ผลไปทิศทางเดียวกับคุณภาพการแปลในภาพรวม กล่าวคือ

1) ข้อผิดพลาดด้านหน่วยคำ (Morphology errors) ระบบ NMT มีข้อผิดพลาดด้านหน่วยคำน้อยกว่าระบบ PBMT อยู่ที่ 19%

2) ข้อผิดพลาดด้านคำศัพท์ (Lexical errors) ระบบ NMT มีข้อผิดพลาดด้านคำศัพท์น้อยกว่าระบบ PBMT อยู่ที่ 17%

3) ข้อผิดพลาดด้านการเรียงลำดับคำ (Word order errors) ระบบ NMT มีข้อผิดพลาดด้านการเรียงลำดับคำน้อยกว่าระบบ PBMT อยู่ที่ 50%

Wu et al. (2017) เปรียบเทียบความสามารถในการแปลระหว่าง PBMT, GNMT และมนุษย์ โดยศึกษาการแปลระบบ NMT ของ Google Translate โดยเฉพาะเพื่อเปรียบเทียบความสามารถกับการแปลของมนุษย์ มีการประเมินผลการแปลโดยผู้เชี่ยวชาญแบ่งออกเป็น 6 ระดับ ระดับ 0 คือแปลไม่รู้เรื่อง ระดับ 6 คือแปลได้อย่างสมบูรณ์ ภาษาที่นำมาเป็นเป้าหมายในการแปลประกอบไปด้วย ภาษาอังกฤษ ภาษาสเปน ภาษาฝรั่งเศส และภาษาจีน โดยเป็นการแปลจากภาษาอังกฤษเป็นภาษาสเปน ภาษาฝรั่งเศส และภาษาจีน และศึกษาการแปลจากภาษาสเปน ภาษาฝรั่งเศส และภาษาจีนเป็นภาษาอังกฤษ ประโยคในการแปลมีจำนวน 500 ประโยคสุ่มมาจาก Wikipedia และเว็บไซต์ข่าว ผลการศึกษาปรากฏตามตารางที่ 1 ด้านล่างซึ่งแสดงให้เห็นว่าระบบ GNMT สามารถแปลภาษาในภาพรวมมีคุณภาพเกือบเทียบเท่าการแปลของมนุษย์

ตารางที่ 1

Mean of side-by-side scores on production data

	PBMT	GNMT	Human	Relative Improvement
English>Spanish	4.885	5.428	5.504	87%
English>French	4.932	5.295	5.496	64%
English>Chinese	4.035	4.594	4.987	58%
Spanish>English	4.872	5.187	5.372	63%
French>English	5.046	5.343	5.404	83%
Chinese>English	3.694	4.263	4.636	60%

ที่มา Wu et al. (2017, p. 19)

แม้ว่าในงานวิจัยของ Wu et al. (2017) ให้ผลวิจัยที่แสดงถึงความสามารถของ GNMT แต่ในงานของ Le & Schuster (2016, September 27) ซึ่งศึกษาการแปลประโยคจากภาษาจีนเป็นภาษาอังกฤษเป็นหลักให้ผลการศึกษาซึ่งแสดงว่า Google Translate ยังมีข้อผิดพลาดที่พบไม่พบในข้อผิดพลาดของมนุษย์ เช่น การแปลคำไม่ครบ การแปลคำนามเฉพาะผิด การแปลคำศัพท์เทคนิคผิด การแปลประโยคแบบเดี่ยว ๆ โดยไม่ครอบคลุมความหมายของข้อความที่ประกอบกันอยู่

ผลการเปรียบเทียบของระบบการแปลภาษาอัตโนมัติแบบ PBMT กับ NMT แสดงให้เห็นว่าเครื่องจักรมีความสามารถในการแปลใกล้เคียงความสามารถของมนุษย์มากยิ่งขึ้น การที่กุเกิลนำระบบการแปล NMT มาใช้ในแพลตฟอร์มแอปพลิเคชันทั้งในรูปแบบการแปลจากอักษรภาพและเสียงยิ่งทำให้การแปลภาษาสะดวกสบาย ในอีกด้านก็เป็นสิ่งที่นักภาษานำจำเป็นต้องเรียนรู้เพื่อหาแนวทางการใช้ระบบการแปลภาษาอัตโนมัติแบบ NMT ของ Google Translate ให้เป็นประโยชน์ต่อนักแปลและเพิ่มประสิทธิภาพการแปลของมนุษย์ให้ดียิ่งขึ้นในลักษณะการอยู่ร่วมกัน ไม่ใช้การเข้ามาทดแทน

จากงานวิจัยศึกษาเปรียบเทียบความสามารถในการแปลทั้งสองงานข้างต้นพบว่าภาษาเป้าหมายที่นำมาเป็นภาษาในการแปลเป็นภาษาพูดและภาษาเขียน โดยงานวิจัยของ Bentivogli et al. (2016) เป็นการศึกษาการแปลภาษาพูด ไม่รวมการศึกษาความสามารถของมนุษย์ในการแปล ส่วนงานของ Wu et al. (2017) ศึกษาการแปลภาษาเขียนระหว่าง PBMT, GMNT และมนุษย์ ผลจากงานวิจัยได้ข้อสรุปคุณภาพการแปลของ Google Translate ในภาพกว้างไม่ได้ลงรายละเอียดว่าชนิดคำหรือกลุ่มคำหรือสำนวน รวมถึงโครงสร้างประโยคแบบใดที่ Google Translate ยังไม่สามารถแปลได้ดีเทียบเท่ากับมนุษย์ทั้งการแปลในระดับคำและระดับประโยค เมื่อคำกลุ่มคำหรือสำนวนนั้นๆ ไปปรากฏ

จากข้างต้นงานวิจัยนี้สรุปเป็นแนวทางการศึกษาได้ว่า การแปลของระบบการแปลภาษาอัตโนมัติที่ผ่านมาในระบบ PBMT นั้นเป็นการแปลแบบเปรียบเทียบคำที่ตรงกันของสองภาษา ซึ่งการแปลเช่นนี้เป็นการแปลรูปแบบหนึ่งที่มีมนุษย์ใช้กันเรียกว่าการแปลแบบตรงตัว คือการแปลให้มีความหมายถูกต้องตรงตามต้นฉบับมากที่สุดคงเนื้อหาไว้ให้ครบถ้วนโดยไม่จำเป็นต้องทำความเข้าใจความหมายก่อน ซึ่งการแปลแบบนี้ของการแปลภาษาอัตโนมัติหรือ Machine Translation เป็นสิ่งที่นักภาษาศาสตร์เห็นว่าไม่สามารถจะแปลได้เทียบเท่าการแปลของมนุษย์ ทั้งนี้หลังจากที่กุเกิลนำศาสตร์ปัญญาประดิษฐ์มาใช้ในการแปลอัตโนมัติในปี ค.ศ. 2016 ความเชื่อเช่นนี้ก็ถูกสั่นคลอน โดยมีการคาดการณ์ว่างานมากกว่า 50% ของนักแปลและล่ามจะถูก Machine Translation หรือเครื่องจักรช่วยแปลแย่งงานไป (Marr, August, 2018) อีกทั้งระบบ NMT เป็นระบบที่โปรแกรมเมอร์พัฒนาให้คอมพิวเตอร์มีความสามารถในการเรียนรู้จากประสบการณ์คล้ายๆ กับสมองของมนุษย์ ผลการวิจัยก่อนหน้านี้แสดงให้เห็นถึงความสามารถในการแปลของระบบ NMT ที่มีความสามารถใกล้เคียงความสามารถของมนุษย์มากยิ่งขึ้น (Bentivogli et al., 2016; Wu et al., 2017) การแปลในระดับให้เกิดความเท่าเทียมกันแบบ Dynamic equivalence ที่จำเป็นต้องเข้าใจภาษารวมถึงวัฒนธรรมด้วยนั้น อาจเกิดขึ้นได้กับเครื่องจักรช่วยแปลในอนาคตอันใกล้

3. สมมติฐานงานวิจัย

Google Translate ระบบ NMT มีแนวโน้มไม่สามารถแปลคำ กลุ่มคำ สำนวน และประโยคที่ใช้ในความหมายระดับข้อความได้อย่างถูกต้องและเหมาะสม

4. วิธีดำเนินการวิจัย

4.1 แหล่งข้อมูล ประกอบไปด้วย 4 กลุ่ม คือ

1) คำ ประกอบไปด้วยคำประสมคำนามและคำประสมคำกริยาจำนวน 136 คำ นำมาจากเอกสารคำสอนรายวิชาโครงสร้างภาษาญี่ปุ่น (ณัฐจิรา ทับทิม, 2559) โดยใช้วิธีการเลือกแบบเจาะจง คำประสมคำนาม จำนวน 36 คำ ประกอบไปด้วยคำประสมคำนามประเภท NsubV (คำนามเป็นประธานคำกริยา) เช่น zaikogire (在庫切れ) NobjV (คำนามเป็นกรรมของคำกริยา) เช่น sakenomi (酒飲み) และ NmodiV (คำนามเป็นส่วนเติมเต็มความหมายของคำกริยา) เช่น tegaki (手書き) จำนวนละ 12 คำ คำประสมคำกริยาจำนวน 100 คำ ประกอบไปด้วยคำประสมคำกริยาดังต่อไปนี้ประเภทละ 5 คำ ได้แก่ v-出す v-合わせる v-取る v-上がる v-上げる v-付ける v-込む v-換える v-替える v-返す v-止める v-起こす v-込める v-入れる v-立てる v-回る v-払う v-渡す v-出る v-受ける

2) กลุ่มคำ ประกอบไปด้วยคำปรากฏร่วมโดยคำปรากฏร่วมหมายถึงคำปรากฏร่วมคือคำที่เกิดจากการรวมกันของคำตั้งแต่ 2 คำขึ้นไป เป็นคำที่มักปรากฏร่วมกัน (Strafella et al., 2012) ประเภทคำปรากฏร่วมที่ใช้ในงานวิจัยนี้ คือ คำนาม+คำกริยา ใช้วิธีเลือกแบบเจาะจงโดยคำนามเป็นคำที่มีความถี่ในการใช้สูง จำนวน 100 คำ อ้างอิงมาจากผลการวิจัยของณัฐจิรา ทับทิม (2561)

3) สำนวน ประกอบไปด้วยสำนวนที่มีคำนามอวัยวะเป็นส่วนประกอบจำนวน 100 สำนวน คำนามอวัยวะที่เป็นส่วนประกอบ 10 คำ คือ atama (頭) me (目) kuchi (口) hana (鼻) mimi (耳) te (手) ki (気) ashi (足) kokoro (心) kubi (首) คำละ 10 สำนวน นำมาจากเอกสารคำสอนรายวิชาโครงสร้างภาษาญี่ปุ่น (ณัฐจิรา ทับทิม, 2559) โดยใช้วิธีการเลือกแบบเจาะจง

4) ประโยค ประกอบไปด้วยประโยคที่ประกอบจากรูปไวยากรณ์ระดับ N4 จำนวน 20 รูป 100 ประโยค นำมาจากข้อสอบวัดระดับความสามารถทางภาษาญี่ปุ่นระดับ 3 (เก่า) ตั้งแต่ปี 1991-2008 (ยกเว้นปี 1999) รวมจำนวน 17 ปี โดยใช้วิธีการเลือกแบบเจาะจง

ทั้งนี้รายการภาษาทั้ง 4 รูปแบบข้างต้นเป็นกลุ่มภาษาที่ไม่สามารถแปลความหมายของคำ กลุ่มคำ สำนวน และประโยคจากความหมายประจำคำหรือความหมายของประโยคเดียวได้

จำต้องอาศัยความรู้เรื่องภาษา บริบท การนำไปใช้ และวัฒนธรรมมาเป็นองค์ประกอบในการแปลความหมาย จึงเป็นรายการภาษาที่มีความหมายในระดับข้อความซึ่งมีความหมายที่สอดคล้องกับวัตถุประสงค์ของงานวิจัย

4.2 เงื่อนไขการสืบค้นข้อมูล

ข้อมูลแต่ละรายการข้างต้นกำหนดเงื่อนไขการสืบค้นคำแปลจาก Google Translate ระบบ NMT ดังนี้

1) สืบค้นด้วยประเภทอักษรตามที่ได้มาจากแหล่งข้อมูลแต่ละแหล่ง เช่น กรณีสำนวน 頭に来る สืบค้นด้วยอักษรคันจิ (頭にくる หรือ あたまに来る หรือ あたまにくる อยู่นอกเหนือเป้าหมายการวิเคราะห์) หรือกรณีคำในประโยค เช่น やさしい สืบค้นด้วยอักษรประเภทฮิรางานะ (優しい หรือ 易しい อยู่นอกเหนือเป้าหมายการวิเคราะห์)

2) สืบค้นรายการข้อมูลประเภทประโยค 2 รูปแบบ คือ รูปแบบที่ใส่ kuten (句点) (。) และรูปแบบที่ไม่ใส่ (พบประโยคที่มีคำแปลแตกต่างกันจากการใส่หรือไม่ใส่ kuten (句点) จำนวน 35 ประโยค โดยมีเพียง 5 ประโยคจากจำนวน 35 ประโยคที่ให้ผลการแปลแตกต่างกัน ทั้งนี้ การนับผลที่ต่างกันของ 5 ประโยค เลื่อนนับผลการแปลที่เป็นที่ยอมรับและที่ยอมรับในระดับปานกลาง)

3) สืบค้นระหว่างเดือนตุลาคมถึงพฤศจิกายน 2561 ผลการศึกษาความสามารถในการแปลความหมายระดับข้อความของ Google Translate ระบบ NMT จากภาษาญี่ปุ่นเป็นภาษาไทย จึงเป็นผลที่สะท้อนความสามารถ ณ ช่วงเวลาที่ทำการศึกษา

4.3 เครื่องมือที่ใช้ในงานวิจัย คือ แบบตรวจสอบการแปลของ Google Translate ระบบ NMT ที่ระบุคำ สำนวน กลุ่มคำ และประโยค จำนวน 436 รายการ

4.4 ขั้นตอนการวิจัย

1) คัดคำ กลุ่มคำ สำนวน และประโยคที่มีความหมายในลักษณะความหมายระดับข้อความ จำนวน 436 รายการ

2) นำคำ กลุ่มคำ สำนวน และประโยคหาคำแปลภาษาไทยจาก Google Translate ระบบ NMT ในแพลตฟอร์ม เว็บไซต์ translate.google.com โดยบันทึกวันเดือนปีที่ค้นหาคำแต่ละตัวอย่าง

3) ส่งแบบตรวจสอบการแปลของ Google Translate ระบบ NMT ให้ผู้ทรงคุณวุฒิที่มีประสบการณ์การเรียนรู้ที่ประเทศญี่ปุ่น มีประสบการณ์การทำงานแปล และงานล่ามภาษาญี่ปุ่นชาวไทย จำนวน 2 คน และคนญี่ปุ่นที่เรียนจบสาขาวิชาภาษาไทยระดับปริญญาเอก และมีประสบการณ์ทำงานสอนที่ประเทศไทย และใช้ภาษาไทยได้เทียบเท่าเจ้าของภาษา จำนวน 1 คน

4) นำผลการประเมินการแปลของผู้ทรงคุณวุฒิทั้ง 3 คน มาหาค่าเฉลี่ยเพื่อวัดความสามารถในการแปลของ Google Translate ระบบ NMT และวิเคราะห์ผลการศึกษา

4.5 เกณฑ์การประเมินการแปลโดยผู้ทรงคุณวุฒิ

เหมาะสม หมายถึง แปลความหมายได้ถูกต้อง สอดคล้องกับความหมายของคำ กลุ่มคำ สำนวน และประโยคในภาษาต้นทาง (ภาษาญี่ปุ่น) เป็นคำแปลที่สามารถสื่อสารให้อีกฝ่ายเข้าใจได้อย่างถูกต้องตามความหมายของภาษาต้นทาง

ไม่เหมาะสม หมายถึง แปลความหมายไม่ถูกต้อง ไม่สอดคล้องกับความหมายของคำ กลุ่มคำ สำนวน และประโยคในภาษาต้นทาง (ภาษาญี่ปุ่น) เป็นคำแปลที่ไม่สามารถสื่อสารให้อีกฝ่ายเข้าใจได้อย่างถูกต้องตามความหมายของภาษาต้นทาง

4.6 การวิเคราะห์ข้อมูลการแปล

ใช้ค่าเฉลี่ยและใช้เกณฑ์การประเมินตามช่วงค่าเฉลี่ยโดยใช้ช่วงคะแนนจากพิสัย (Intervals from the Range) และมีเกณฑ์การแปลความหมายดังนี้

0.68 – 1.00 คำแปลภาษาไทยของ Google Translate ระบบ NMT เป็นที่ยอมรับ

0.34 – 0.67 คำแปลภาษาไทยของ Google Translate ระบบ NMT เป็นที่ยอมรับในระดับปานกลาง

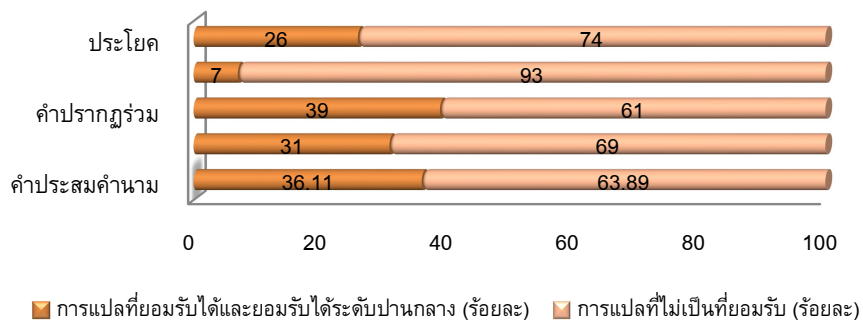
0.00 – 0.33 คำแปลภาษาไทยของ Google Translate ระบบ NMT ไม่เป็นที่ยอมรับ

5. ผลการวิจัยและการอภิปรายผล

ผลความสามารถในการแปลของ Google Translate ระบบ NMT ภาพรวม

แผนภาพที่ 1

ความสามารถในการแปลของ Google Translate ระบบ NMT

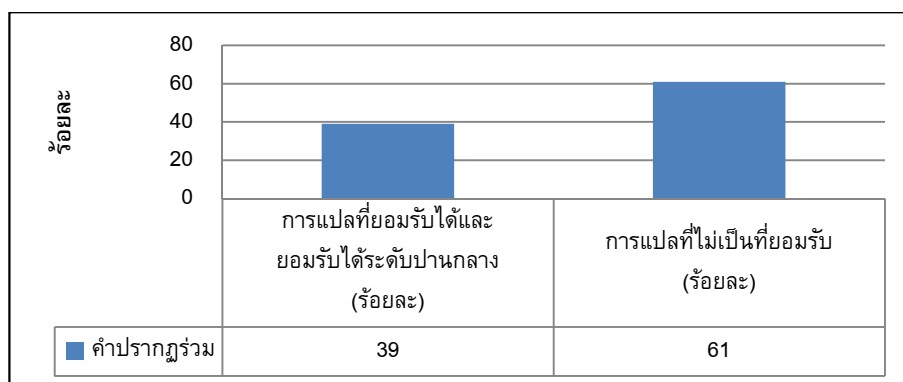


จากแผนภาพที่ 1 จะเห็นได้ว่า Google Translate ระบบ NMT สามารถแปลความหมายของคำ สำนวน และประโยคในระดับยอมรับได้และยอมรับได้ปานกลางอยู่ที่ 116 รายการ คิดเป็นร้อยละ 26.61 และสามารถแปลในระดับที่ไม่เป็นที่ยอมรับอยู่ที่ 320 รายการ คิดเป็นร้อยละ 73.39 โดยสามารถแปลในระดับยอมรับได้และยอมรับได้ระดับปานกลางแยกออกได้เป็นคำประสมคำนาม ร้อยละ 36.11 คำประสมคำกริยา ร้อยละ 31 คำปรากฏรวม ร้อยละ 39 สำนวน ร้อยละ 7 และประโยค ร้อยละ 26

กลุ่มภาษาที่ Google Translate ระบบ NMT แปลได้ดีที่สุดคือกลุ่มภาษาประเภทคำปรากฏรวม จำนวนคำที่แปลได้ คือ 39 คำ คิดเป็นร้อยละ 39 จากจำนวน 100 คำ ตามแผนภาพที่ 2

แผนภาพที่ 2

ผลการแปลคำปรากฏรวม



ยกตัวอย่างคำและคำแปลได้ดังนี้ (ในวงเล็บคือคำแปลภาษาไทยที่ Google Translate ระบบ NMT แปล) imi o motsu (意味を持つ) (มีความหมาย) riyū o tsukeru (理由をつける) (ให้เหตุผล) eikyō o ataeru (影響を与える) (มีผลต่อ) kōka ga deru (効果が出る) (มีประสิทธิภาพ) jōken o mitasu (条件を満たす) (ปฏิบัติตามเงื่อนไข) jōtai ni ochiiru (状態に陥る) (ตกอยู่ในสถานะ) kinō ga teikasuru (機能が低下する) (ฟังก์ชันแย่งลง) jōkyō o haakusuru (状況を把握する) (ทำความเข้าใจสถานการณ์) mokuhyō o tasseisuru (目標を達成する) (บรรลุเป้าหมาย) sa-bisu o teikyōsuru (サービスを提供する) (ให้บริการ)

ทั้งนี้ผลการศึกษาการแปลคำปรากฏรวมข้างต้นมีแนวโน้มสอดคล้องกับลักษณะของ Google Translate ระบบ NMT ที่ถูกสร้างมาด้วยระบบคลังข้อมูลรูปแบบ Corpus-based machine translation ซึ่งอาศัยการรวบรวมข้อมูลการแปลเป็นภาษาต่างๆ จัดเก็บเป็นคลังข้อมูล

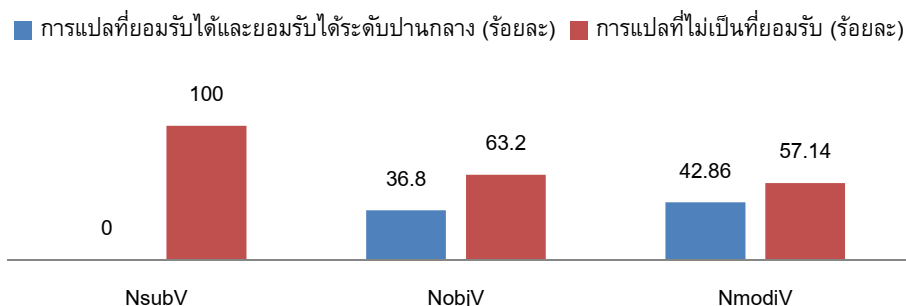
ภาษาขนาดใหญ่ที่มีคำศัพท์รวมถึงคำปรากฏรวมที่เกิดขึ้นบ่อยครั้ง โดยธรรมชาติของคำปรากฏรวมเป็นคำกลุ่มที่สามารถแปลความหมายได้จากความหมายประจำคำแต่ละคำและเมื่อพิจารณากลุ่มคำปรากฏรวมที่สามารถแปลได้พบว่ามีแนวโน้มเป็นกลุ่มคำนามที่ปรากฏรวมกับคำกริยาที่มีความถี่ในการใช้สูงในกลุ่มคำปรากฏรวมของคำนามนั้นๆ (จากการสืบค้นจากคลังข้อมูลภาษา NINJAL-LWP for TWC (<http://nlt.tsukuba.lagoinst.info/>) ในงานวิจัยของณัฐริธา ทับทิม, 2561) เช่นคู่คำนามที่มีความหมายคล้ายคำว่า jōtai และ jōkyō (状態・状況) ซึ่งเป็นคำนามที่มีความถี่การใช้จริงเป็นอันดับ 11 และ 12 จาก 100 คำนาม (ณัฐริธา ทับทิม, 2561) คำปรากฏรวม jōtai ni ochiiru (状態に陥る) (ตกอยู่ในสถานะ) เป็นคำปรากฏรวมที่มีความถี่สูงสุดของคำปรากฏรวมของคำนาม jōtai (状態) ส่วนคำปรากฏรวม jōkyō o haakusuru (状況を把握する) เป็นคำปรากฏรวมที่มีความถี่สูงสุดของคำปรากฏรวมของคำนาม jōkyō (状況)

คำปรากฏรวมอื่นๆ ในจำนวน 39 คำ ส่วนใหญ่เป็นคำปรากฏรวมที่มีความถี่สูงในกลุ่มของการใช้ของคำนามนั้นๆ เช่นเดียวกับกรณีของคำนาม jōtai และ jōkyō (状態・状況) ยกเว้นบางคำ เช่น riyū o tsukeru (理由をつける) (ให้เหตุผล) kōka ga deru (効果が出る) ที่ความถี่การใช้ของคำปรากฏรวมมีความถี่การใช้เป็นอันดับ 2 ในการใช้แบบคำปรากฏรวมของคำนาม riyū (理由) และ kōka (効果)

ส่วนกลุ่มคำปรากฏรวมที่ Google Translate ระบบ NMT ยังแปลได้ในระดับยอมรับไม่ได้มีแนวโน้มส่วนใหญ่เป็นคำปรากฏรวมที่มีความถี่ไม่ใช่อันดับสูงสุดของคำนามนั้นๆ จำนวน 33 คำ (คิดเป็นร้อยละ 54) จากจำนวนทั้งหมดที่แปลไม่ได้ คือ 61 คำ เช่น kankyō ni hairyo (環境に配慮する) (ดูแลสิ่งแวดล้อม) shien o ukeru (支援を受ける) (ต้องการรับความช่วยเหลือ) anzen o kakuho (安全を確保する) (เพื่อความปลอดภัย) bunka ni nedoku (文化に根付く) (รากในวัฒนธรรม) jiyū o ubau (自由を奪う) (ปลดปล่อยอิสรภาพ) shiryō o toriyoseru (資料を取り寄せる) (สต็อคส่งซื้อสินค้า) sentaku o semaru (選択を迫る) (เพื่อให้เป็นทางเลือก) omoi o yoseru (思いを寄せる) (เพื่อนำมาคิด) pointo o matomeru (ポイントをまとめる) (เก็บคะแนน) jiken ga hakkakusuru (事件が発覚する) (มีการค้นพบกรณี)

แผนภาพที่ 3

ผลการแปลคำประสมคำนาม



จากตารางที่ 1 และแผนภาพที่ 3 ข้างต้น Google Translate ระบบ NMT สามารถแปลกลุ่มคำประเภทคำประสมคำนามได้ 13 คำ คิดเป็นร้อยละ 36.11 แปลไม่ได้ 23 คำ คิดเป็นร้อยละ 63.89 แยกออกเป็นคำประสมคำนามประเภท NsubV มีการแปลในระดับไม่เป็นที่ยอมรับ ร้อยละ 100 คำประสมคำนามประเภท NobjV มีการแปลในระดับยอมรับได้ ร้อยละ 36.8 และการแปลในระดับไม่เป็นที่ยอมรับร้อยละ 63.2 และคำประสมคำนามประเภท NmodiV มีการแปลในระดับยอมรับได้ ร้อยละ 42.86 และการแปลในระดับไม่เป็นที่ยอมรับร้อยละ 57.14

คำที่แปลได้ จำนวน 13 คำ เช่น kankiri (缶切り) (ที่เปิดกระป๋อง) sennuki (栓抜き) (ที่เปิดขวด) hitogoroshi (人殺し) (ฆาตกร) usotsuki (嘘つき) (คนโกหก) kubikazari (首飾り) (สร้อยคอ) bunkatsubarai (分割払い) (ผ่อนชำระ) maegari (前借り) (เงินกู้ยืมล่วงหน้า) taue (田植え) (ปลูกข้าว) mizuarai (水洗い) (ซักด้วยน้ำ)

จากการพิจารณากลุ่มคำนามประสม 13 คำที่ Google Translate ระบบ NMT แปลได้ ความหมายของกลุ่มคำประสมคำนามส่วนใหญ่เป็นคำที่ใช้เป็นชื่อเรียกสิ่งของที่เป็นอุปกรณ์ หรือสิ่งของทั่วไป (ที่เปิดกระป๋อง ที่เปิดขวด สร้อยคอ เงินกู้ยืมล่วงหน้า) ชื่อเรียกคน (คนโกหก) หรือชื่อเรียกกิจกรรม ((การ) ผ่อนชำระ (การ) ปลูกข้าว (การ) ซักด้วยน้ำ) และพบการใช้จริงทุกคำจากการสำรวจจากคลังข้อมูลภาษา NINJAL-LWP for TWC ในขณะที่คำประสมคำนามจำนวน 23 คำ เป็นคำที่ Google Translate ระบบ NMT แปลในระดับที่ยอมรับไม่ได้ โดยจำนวน 10 คำ ได้แก่ zaikokire (在庫切れ) 、kitaihazure (期待外れ) 、bōshikake (帽子掛け) 、zōtsukai (象使い) 、mēshiire (名刺入れ) 、nezumitorie (鼠捕り) 、gomiherashi (ゴミ減らし) 、kingyosukui (金魚掬い) 、kōrisuberi (氷滑り) 、takunomi (宅飲み) เป็นคำที่ไม่พบการใช้จริงในคลังข้อมูลภาษา NINJAL-LWP for TWC รวมถึงคำประสมคำนามบางคำพบการใช้จริงใน

ระดับความถี่ไม่เกิน 100 ครั้ง (จำนวน 3 คำ ได้แก่ 茶飲み、骨休み、足踏み) หรือแม้ว่าคำประสมคำนามบางคำจะมีความถี่ในการใช้ค่อนข้างสูง เช่น higaeri (日帰り)、kandume (缶詰め)、hidarikiki (左利き)、sakenomi (酒飲み)、yunomi (湯飲み)、tamagoyaki (卵焼き)、shokkierai (食器洗い)、sekibara (咳払い) แต่ Google Translate ระบบ NMT ก็ไม่สามารถแปลในระดับที่ยอมรับได้

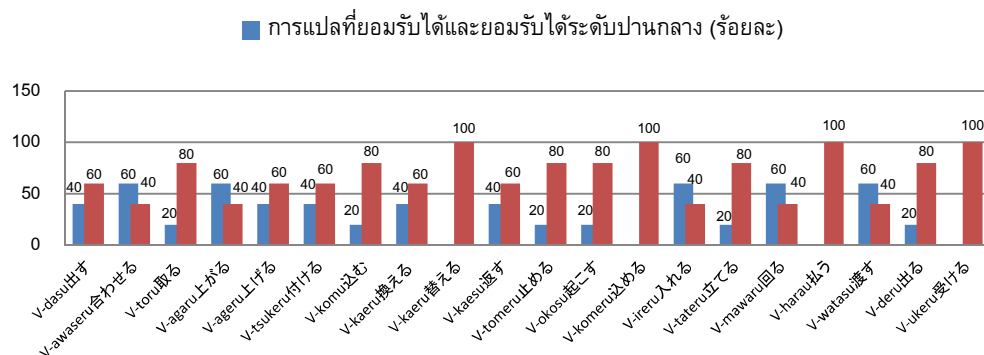
เนื่องจากคำประสมคำนามข้างต้นส่วนใหญ่เป็นกลุ่มคำที่มีความหมายข้อความหรือต้องอาศัยบริบทของข้อความ วัฒนธรรมของการใช้คำเป็นองค์ประกอบในการเข้าใจความหมาย ความหมายของคำประสมคำนามส่วนใหญ่กลุ่มนี้จึงอยู่ในกลุ่มความหมายโวหารภาพพจน์ประเภทนามนัยทั้งนามนัยเชิงเวลา เชิงพื้นที่ เชิงเหตุผล หรือลักษณะเชิงสัญลักษณ์บางส่วนมาแทนสิ่งหนึ่งทั้งหมด (ณัฐริธา ทับทิม, 2559ก) ความหมายที่เหมาะสมของคำ เช่น sekibara (咳払い) (Google Translate ระบบ NMT แปลเป็น “เพื่อล้างลำคอ” ซึ่งในการแปลควรเลือกใช้คำ “กระแอม” ((การ) ทำเสียงแอมในคอคล้ายไอ เพื่อให้โปร่งคอ ให้เสียงหายเคื้อ เพื่อให้อาเจียรหรือระวัง เป็นต้น)) เป็นความหมายที่เกิดจากลักษณะของการกระทำที่ส่งเสียงไอซึ่งดังหลายครั้งทำให้คนหันมาสนใจ ลักษณะเช่นนี้ถูกนำมาใช้ในความหมายกระแอมเพื่อให้คนอื่นรู้ หรือเรียกความสนใจ เป็นต้น เป็นความหมายที่เกิดจากการใช้จริงจนมีความหมายขยายในความหมาย “กระแอม” ซึ่ง Google Translate ระบบ NMT ยังไม่สามารถแปลความหมายลักษณะนี้ของคำประสมคำนามได้ยกเว้นคำที่มีการใช้จริงค่อนข้างมากและพบคำแปลภาษาไทยที่ใช้แทนกันเช่น คำประสมคำนาม 13 คำ ข้างต้นที่อยู่ในกลุ่มการแปลในระดับยอมรับได้

จากข้างต้นจึงสรุปได้ถึงความสามารถในการแปลของ Google Translate ระบบ NMT ว่ายังไม่สามารถแปลความหมายระดับข้อความในกลุ่มโวหารภาพพจน์ได้ดีเท่ากับมนุษย์ที่มีความสามารถในการสืบค้นวัฒนธรรมของการใช้คำ การคิดวิเคราะห์โครงสร้างเชิงลึกของคำ

กรณีการแปลคำประสมอีกชนิดหนึ่งคือคำประสมคำกริยา ในการแปลกลุ่มคำชนิดนี้ก็พบผลเช่นเดียวกับคำประสมคำนามคือความสามารถในการแปลของ Google Translate ระบบ NMT ยังอยู่ในระดับไม่เป็นที่ยอมรับถึงร้อยละ 69 มีเพียงร้อยละ 31 เท่านั้นที่สามารถแปลในระดับที่ยอมรับได้ตามแผนภาพที่ 1 โดยแจกแจงรายละเอียดได้ตามแผนภาพที่ 4

แผนภาพที่ 4

ผลการแปลคำประสมคำกริยา



ทั้งนี้แม้ว่าความหมายของคำประสมคำกริยาจะไม่ใช่ความหมายกลุ่มโวหารภาพพจน์ แต่การแปลความหมายของคำกริยาเหล่านี้ต้องอาศัยความรู้เรื่องโครงสร้างคำที่มาประสมกัน ซึ่ง Google Translate ระบบ NMT มีแนวโน้มยังขาดข้อมูลอย่างเพียงพอที่จะใช้เพื่อการเรียนรู้ด้วยตนเองในเรื่องโครงสร้างคำประสมซึ่งสอดคล้องกับลักษณะของเครื่องจักรช่วยแปลในรูปแบบ Corpus-based machine translation ของ Google Translate ระบบ NMT ที่ต้องอาศัยข้อมูลขนาดใหญ่ เนื่องจากหากมีข้อมูลการแปลคำประสมคำกริยาจากภาษาญี่ปุ่นเป็นภาษาไทยมากพอ Google Translate ระบบ NMT ก็อาจสามารถเรียนรู้ด้วยตนเองเพื่อแปลความหมายคำออกมาได้ เช่นในกรณีคำประสมคำกริยา กลุ่ม –awaseru (-合わせる) โดย Google Translate ระบบ NMT สามารถแปลอยู่ในระดับยอมรับได้ร้อยละ 60 หรือ 3 คำจาก 5 คำ คือ kumiawaseru (組み合わせる) (รวมกัน) toiwaseru (問い合わせる) (สอบถาม) และ miawase (見合わせる) (ชะลอ) ส่วนคำประสมคำกริยาอีก 2 คำที่มีการแปลอยู่ในระดับไม่เป็นที่ยอมรับ คือ uchiawaseru (打ち合わせる) (เพื่อตอบสนอง) และ kuriawaseru (繰り合わせる) (ทำซ้ำ) ทั้งนี้ผลการแปลสัมพันธ์กับระดับความถี่ของคำที่ปรากฏในคลังข้อมูลภาษา NINJAL-LWP for TWC ซึ่ง kumiawaseru (組み合わせる) toiwaseru (問い合わせる) และ miawaseru (見合わせる) มีความถี่การใช้สูงกว่าคำประสมคำกริยาอื่นในกลุ่มเดียวกัน คือ 28,804 9,829 และ 2,396 ครั้ง (ส่วนอีก 2 คำมีความถี่ 804 และ 57 ตามลำดับ)

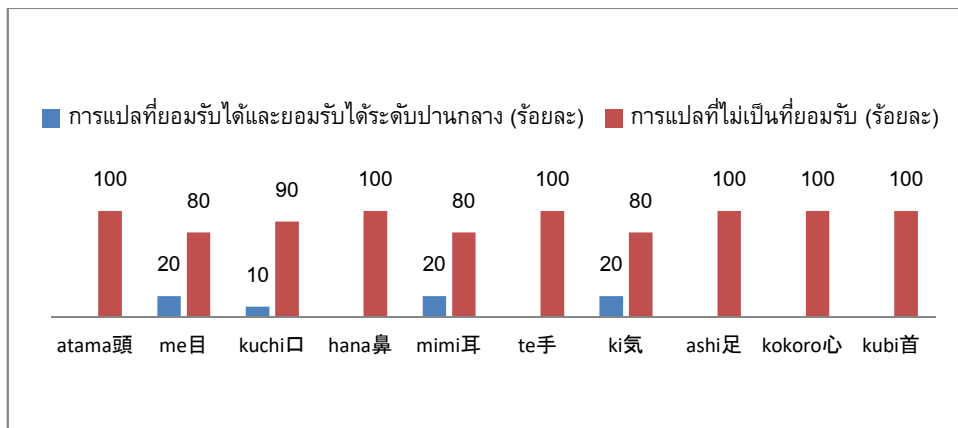
ตัวอย่าง –tomeru (-止める) โดย Google Translate ระบบ NMT สามารถแปลอยู่ในระดับยอมรับได้ร้อยละ 20 หรือเพียง 1 คำ คือ uketomeru (受け止める) (ยอมรับ) ส่วนคำประสมคำกริยาอีก 4 คำ ในกลุ่มเดียวกันที่มีการแปลอยู่ในระดับไม่เป็นที่ยอมรับ คือ tsukitomeru (突き止める) (ตั้งอยู่) hikitomeru (引き止める) (ถอยกลับ), yobitomeru (呼び止める) (เมื่อต้องการหยุด) sashitomeru (差し止める) (เมื่อต้องการหยุด) มีคำแปลอยู่ในระดับที่ยอมรับไม่ได้ ซึ่ง uketomeru

(受け止める) มีความถี่การใช้สูงกว่าคำประสมคำกริยาอื่นในกลุ่มเดียวกันคือ 31,022 ครั้ง (ส่วนอีก 4 คำมีความถี่ 3,574 1,835 1,205 และ 651 ตามลำดับ)

กลุ่มคำที่ Google Translate ระบบ NMT แปลได้น้อยมากหรือกล่าวได้ว่าแทบไม่มีความสามารถในการแปล คือ กลุ่มคำสำนวน ตามแผนภาพที่ 1 ข้างต้นและแผนภาพที่ 5

แผนภาพที่ 5

ผลการแปลสำนวน

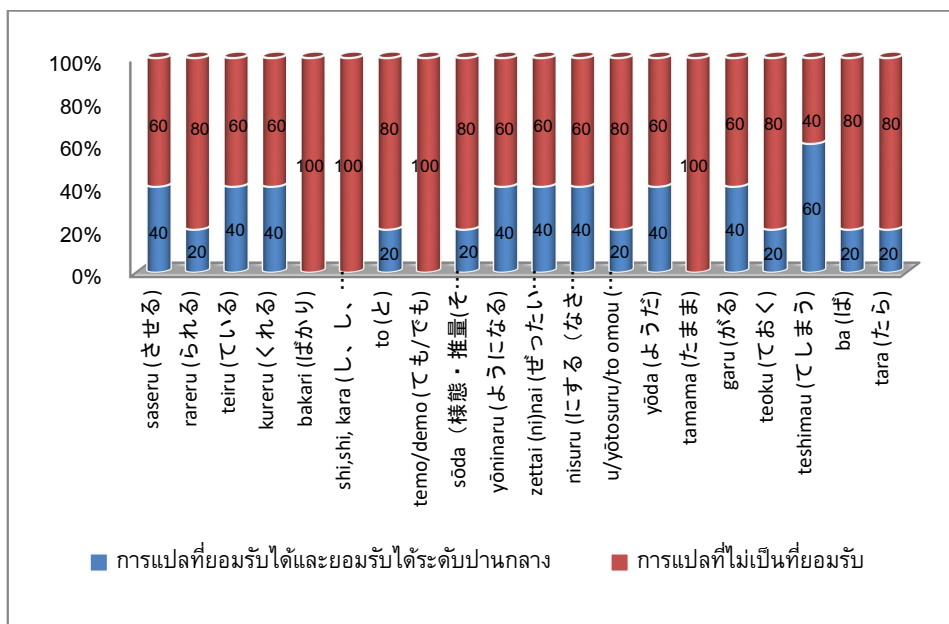


กลุ่มคำสำนวนมีร้อยละการแปลในระดับที่ยอมรับได้เพียงร้อยละ 7 โดยกลุ่มคำสำนวนเป็นกลุ่มคำที่ไม่สามารถคาดเดาความหมายจากความหมายของคำแต่ละคำที่มารวมกัน และไม่มีกรณการรวมกันโดยอาศัยโครงสร้างของคำเหมือนคำประสมคำนามหรือคำประสมคำกริยาซึ่งหากเข้าใจโครงสร้างที่รวมกันก็อาจเชื่อมโยงถึงการแปลความหมายของคำ สำนวนที่ Google Translate ระบบ NMT สามารถแปลได้ร้อยละ 7 หรือ 7 คำ คือ mimi o katamukeru (耳を傾ける) (ฟัง) kuchi ga omoi (口が重い) (ปากหนัก) ki ga au (気が合う) (เข้ากันได้ดี) mimi o sumasu (耳を澄ます) (ฟังอย่างระมัดระวัง) me ga sameru (目が覚める) (ตื่นขึ้นมา) ki o tsukeru (気をつける) (ต้องระวัง), me o samasu (目を覚ます) (ตื่นขึ้นมา)

กรณีของประโยค Google Translate ระบบ NMT ยังไม่สามารถแปลประโยคได้ดี เมื่อวิเคราะห์ความสามารถในการแปลโครงสร้างของประโยคที่ประกอบไปด้วยรูปไวยากรณ์ 20 รูป รูปละ 5 ประโยค นำเสนอผลการแปลประโยคภาพรวมตามแผนภาพที่ 1 ได้ คือ Google Translate ระบบ NMT สามารถแปลในระดับยอมรับได้และยอมรับได้ระดับปานกลางอยู่ที่ร้อยละ 26 และแปลในระดับที่ไม่เป็นที่ยอมรับอยู่ที่ร้อยละ 74 แจกแจงรายละเอียดด้วยแผนภาพที่ 6 ได้ดังนี้

แผนภาพที่ 6

ความสามารถในการแปลของ Google Translate



จากแผนภาพที่ 6 จะเห็นได้ว่าผลการประเมินความเหมาะสมของความหมายจากผู้ทรงคุณวุฒิทั้งสามท่าน ส่วนใหญ่คำแปลภาพรวมของประโยคมีความหมายไม่เป็นที่ยอมรับ โดยประโยคที่มีรูปไวยากรณ์ teshimau (てしまう) เป็นกลุ่มประโยคที่ Google Translate ระบบ NMT แปลได้เหมาะสมมากที่สุดคือร้อยละ 60 หรือจำนวน 3 ประโยคดังนี้

- (1) michi ni mayotteshimaimashita (道に迷ってしまいました)

ฉันหลงทาง

- (2) kusuri o zenbu nondeshimaimashita (薬を全部飲んでしまいました)

ฉันกินยาทั้งหมด

- (3) ashi ni kega o shichatta (足にけがをしちゃった)

ฉันได้รับบาดเจ็บที่เท้าของฉัน

จากประโยคแปลเห็นได้ว่าเป็นการแปลความหมายที่สื่อสารให้กับผู้อ่านได้ ไม่ได้แตกต่างจากความหมายของภาษาต้นทางที่ต้องการสื่อ ทั้งนี้การใช้รูปไวยากรณ์ teshimau (てしまう)

ใช้ใน 2 ลักษณะ คือ การกระทำหรือเหตุการณ์เสร็จสิ้นสมบูรณ์กับแสดงความรู้สึกเสียตายเป็นใจที่ตัดสินใจผิดพลาด (Group Jammassy, 2554, น. 326-327) โดยในส่วนของคำแปลนี้ความหมายการใช้รูปไวยากรณ์ teshimau (てしまう) ทั้ง 3 ประโยคข้างต้นค่อนข้างเหมาะสมกับลักษณะของการใช้ทั้งสองลักษณะ

ประโยคที่ Google Translate ระบบ NMT ไม่สามารถแปลได้คือประโยคที่มีรูปไวยากรณ์ bakari (ばかり) shi shi kara (し、し、から) temo/demo (ても・でも) และรูป tamama (たまま) ปรากฏตามแผนภาพที่ 6 ข้างต้น ตัวอย่างผลการแปลของประโยคที่ประกอบไปด้วยรูปไวยากรณ์ดังกล่าวได้แก่

- (4) uchi no kodomo wa mainichi ju-su bakari nondeimasu

(うちの子供は毎日ジュースばかり飲んでいます)

ลูกของเราดื่มน้ำวันละครั้งเท่านั้น

- (5) sensei wa chucchōshite bakari de, jugyō o suru jikan ga nai

(先生は出張してばかりで、授業をする時間がない)

ครูเป็นเพียงการเดินทางธุรกิจและฉันไม่มีเวลาสอน

จากคำแปลข้างต้นจะเห็นได้ว่านอกจากการแปลความหมายคำที่เป็นส่วนประกอบของประโยคไม่เหมาะสมแล้วการแปลการใช้ของรูปไวยากรณ์ก็ผิดไปจากความหมายที่ควรแปลคือ “มีแต่สิ่งนั้นเท่านั้น ไม่มีสิ่งอื่น” (Group Jammassy, 2554, น. 615)

- (6) mō osoishi, tsukareta kara, massugu uchi ni kaerō

(もう遅いし、疲れたから、まっすぐうちに帰ろう)

มันเหนื่อยและเหนื่อยแล้วก็ไปบ้านกันเถอะ

- (7) otto wa yasashii shi, kakkoi shi, suteki na hito desu

(夫はやさしいし、かっこいいし、すてきな人です。)

สามีของฉันเป็นคนใจดีและเป็นคนดี

รูป shi shi kara (し、し、から) ใช้ในความหมายเพื่อแสดงการเรียงของข้อความที่คล้ายตามกันและแสดงเหตุผล (Group Jammassy, 2554, น. 179-181) จากคำแปลจะเห็นได้ว่าการแปลการใช้รูป shi shi kara (し、し、から) ที่ปรากฏในประโยคไม่ได้เป็นจุดที่ไม่เหมาะสมจุดที่ไม่เหมาะสมคือการแปลความหมายภาพรวมของประโยคและการแปลส่วนประกอบประโยค

ไม่ครบ ประเด็นนี้แตกต่างกับผลการแปลประโยคที่มีรูป bakari (ばかり) ซึ่งมีลักษณะการเปลี่ยนความหมายของรูปไวยากรณ์ไม่เหมาะสม (ความหมายภาษาไทยในวงเล็บข้อ (8) (14) และ (15) เป็นคำแปลกรณีใส่ kuten (句点。)) ตามเงื่อนไขการสืบค้นที่อธิบายไว้ข้างต้น)

(8) anosensei wa byōkini nattemo gakkō o yasumimasendeshita

(あの先生は病気になっても学校を休みませんでした)

ครูคนนั้นไม่ได้ไปโรงเรียนสักวันหนึ่งแม้จะป่วย (ครูคนนั้นไม่ได้ไปโรงเรียนสักวันหนึ่งแม้ว่าจะป่วยก็ตาม)

(9) nankai hon o yondemo, wakaranakatta node, sensei ni kiku kotonishimashita

(何回本を読んでも、分からなかったの、先生に聞くことにしました)

ไม่ก็ครั้งที่ฉันอ่านหนังสือฉันไม่เข้าใจดังนั้นฉันจึงตัดสินใจถามครู

จากคำแปลจะเห็นได้ว่า Google Translate ระบบ NMT สามารถเปลี่ยนความหมายการใช้รูป temo/demo (ても・でも) ซึ่งแสดงเงื่อนไขขัดแย้ง (แม้ว่า...สักก็ครั้ง) ได้เหมาะสม แต่ความหมายภาพรวมของประโยคไม่เหมาะสมจึงทำให้ถูกประเมินในระดับไม่เป็นที่ยอมรับ ซึ่งมีลักษณะข้อผิดพลาดเหมือนกับการแปลประโยคที่มีรูปไวยากรณ์ bakari (ばかり)

(10) uchi no denki o tsuketamama kaisha ni kiteshimaimashita

(うちの電気をつけたまま会社に来てしまいました)

ฉันมาที่ออฟฟิศด้วยไฟไฟฟ้าของฉัน

(11) terebi o tsuketamama neteshimaimashita

(テレビをつけたまま、寝てしまいました)

ฉันหลับไปพร้อมกับทีวี

(12) netamama tabenaide. shitsurei ni mierukara

(寝たまま食べないで。しつれいに見えるから)

อย่ากินในขณะที่นอนหลับ เพราะดูไม่สุภาพ

(13) ame ni nuretamama dato kaze o hikuyo

(雨に濡れたままだとかぜをひくよ)

ฉันจะเป็นหวัดถ้าฉันเปียกฝน

(14) itsumo pajama o kitamama shokuji o suru nowa yokunai yo

(いつもパジャマを着たまま食事をするのはよくないよ)

มันไม่ดีที่จะกินเสมอในขณะที่สวมใส่ชุดนอน (ไม่ควรทานกับชุดนอนเสมอไป)

กรณีการแปลประโยคที่มีรูปไวยากรณ์ tamama (たまま) ที่มีการใช้ในบริบทเรื่องหรือเหตุการณ์อยู่ในสภาพเดิมไม่เปลี่ยนแปลง (Group Jammassy, 2554, น. 684) จากการแปลประโยคทั้งหมด 5 ประโยคข้างต้นมีข้อสังเกตที่ได้แตกต่างจากรูปไวยากรณ์ 3 รูปแรกคือผลการแปลมีทั้งที่ไม่เหมาะสมในการแปลการใช้รูปไวยากรณ์ tamama (たまま) คือประโยค (10) (11) (13) (14) (กรณีประโยคในวงเล็บ) และมีผลการแปลส่วนประกอบประโยครวมถึงความหมายภาพรวมของประโยคไม่เหมาะสมแม้ว่าจะจะเปลี่ยนการใช้รูปไวยากรณ์เหมาะสมก็ตาม (ประโยค (12) และ (14))

จากข้อสังเกตข้างต้นสามารถสรุปแนวโน้มความสามารถในการแปลประโยคของ Google Translate ระบบ NMT เมื่อสังเกตผลการแปลแยกออกเป็นระดับคำ ระดับโครงสร้าง และระดับความหมายภาพรวมพบว่าส่วนใหญ่ Google Translate ระบบ NMT แปลนัยการใช้ของโครงสร้างแต่ละรูปไวยากรณ์ได้เหมาะสมและลักษณะประโยคที่เป็นประโยคความเดียวหรือประโยคความซ้อนไม่ได้เป็นปัจจัยที่ทำให้ความสามารถในการแปลลดลงหรือเพิ่มขึ้น แต่สาเหตุที่ทำให้ผลการประเมินจากผู้ทรงคุณวุฒิทั้งสามคนไม่ผ่านมีแนวโน้มเป็นเพราะการแปลความหมายในระดับคำที่พบข้อผิดพลาดจำนวนมากซึ่งรวมถึงการแปลคำไม่ครบ เช่น ในประโยค (4)-(7) และ (10)-(14) เช่นการแปล ju-su (ジュース) ว่า “น้ำ” การแปล osoi (遅い) ว่า “เหนื่อย” หรือการแปลไม่ครบหรือตกหล่นเช่นประโยค (7)

ตัวอย่างการแปลนัยความหมายการใช้ของบางรูปไวยากรณ์ที่เหมาะสมตามข้อสังเกตข้างต้นเพิ่มเติมเช่น ในกรณีรูปทำให้กระทำ saseru (させる) Google Translate ระบบ NMT สามารถแปลนัยการใช้โครงสร้างของรูปไวยากรณ์ได้ทุกประโยค แม้จะมีผลการแปลความหมายทั้งประโยคผิดร้อยละ 60 ก็ตาม เช่น

(15) Otōto wa mata chichi o okoraseta

(弟はまた父を怒らせた)

พี่ชายของฉันทำให้พ่อของฉันโกรธ (น้องชายทำให้พ่อโกรธอีกครั้ง)

จากประโยคข้างต้นหากพิจารณาความหมายที่แปลทั้งแบบใส่ kuten (句点) และไม่ใช่ จะเห็นว่า Google Translate ระบบ NMT สามารถแปลโครงสร้างของรูปประโยคทำให้กระทำได้ดี เทียบเท่ากันกับประโยคที่แปลได้เหมาะสมในวงเล็บ เนื่องจากประโยคแปลมีนัยความหมาย ตามโครงสร้างรูปไวยากรณ์ “ทำให้กระทำ” แม้ว่าจะแปลคำว่า otōto (弟) ผิดก็ตาม หรือในกรณี รูปถูกกระทำ rareru (られる) Google Translate ระบบ NMT ก็สามารถแปลโครงสร้างรูปประโยค ถูกกระทำได้ดีเช่นประโยค (16) เป็นประโยคที่ถูกประเมินในระดับยอมรับได้ ส่วนประโยค (17) อยู่ในกลุ่มที่ Google Translate ระบบ NMT ยังไม่สามารถแปลในระดับยอมรับได้ แต่เมื่อพิจารณา การแปลในระดับโครงสร้างประโยครูปถูกกระทำจะเห็นได้ว่าเป็นการแปลผ่านโครงสร้างกระทำ (Active voice) ที่สามารถสื่อสารได้แม้ว่าการแปลในระดับคำคือ yoru (夜) จะไม่เหมาะสมก็ตาม

(16) Kono tatemono wa yūmēna kenchikuka niyotte tateraremashita

(このたてものは有名な建築家によって建てられました)

อาคารนี้สร้างโดยสถาปนิกชื่อดัง

(17) Kinō no yoru zutto akachan ni nakarete tsukaremashita.

(きのうの夜ずっと赤ちゃんに泣かれて疲れました)

ฉันเหนื่อยหลังจากที่ลูกร้องไห้ตอนเย็นเมื่อวานนี้

ในการพิจารณาแปลโครงสร้างประโยคที่มีรูปไวยากรณ์อื่นๆ ก็มีแนวโน้มการเปลี่ยนการใช้ของรูปไวยากรณ์ส่วนใหญ่เหมาะสมแม้ว่าผลการประเมินความหมายภาพรวมทั้งประโยคจะอยู่ในระดับไม่เป็นที่ยอมรับถึงร้อยละ 74 ก็ตาม

ผลการวิจัยที่พบว่า Google Translate ระบบ NMT แปลโครงสร้างประโยคหรือลักษณะการใช้รูปไวยากรณ์ได้นั้นให้ผลที่แตกต่างจากงานวิจัยที่ศึกษาการแปลภาษาอังกฤษเป็นภาษาเปอร์เซียจากเครื่องจักรช่วยแปล 3 ชนิด (หนึ่งในนั้นคือ Google Translate) ในงานของ Komeili, Hendavalan & Rahimi (2011) ที่สรุปว่าเครื่องจักรช่วยแปลยังไม่สามารถแปลโครงสร้างประโยค (Syntax) ได้ดี โดยผลจากการเก็บข้อมูลการแปลรูปไวยากรณ์ทั้งหมด 20 รูปจากการแปลจากภาษาญี่ปุ่นเป็นภาษาไทยให้แนวโน้มว่า Google Translate ระบบ NMT สามารถรักษาการแปลตามหลักการใช้รูปไวยากรณ์แต่ละรูปได้ ถึงแม้ว่าผลการแปลความหมายของคำที่เป็นส่วนประกอบในประโยคจะผิดซึ่งทำให้ผลการประเมินความหมายภาพรวมอยู่ในระดับที่ยอมรับไม่ได้ไปด้วยก็ตาม ผลการวิจัยยังสอดคล้องกับ Le & Schuster ที่พบว่า Google Translate ยังมีกรณีแปลไม่ครบ และแปลประโยคตกหล่น อีกทั้งยังมีปัญหาการแปลคำที่มีหลายความหมาย

ผลการวิจัยข้างต้นสนับสนุนสมมติฐานที่ตั้งไว้ว่า Google Translate ระบบ NMT มีแนวโน้มไม่สามารถแปลคำ กลุ่มคำ สำนวน และประโยคที่ถูกใช้ในความหมายระดับข้อความได้อย่างถูกต้องและเหมาะสมโดยผลการวิจัยได้นำไปปรับใช้เป็นข้อสังเกตหนึ่งในการสอนเนื้อหาทรัพยากรการแปลและการล่าม (อธิบายเกี่ยวกับลักษณะของพจนานุกรมทั่วไป พจนานุกรมศัพท์เฉพาะ เครื่องจักรช่วยแปลประเภทต่าง ๆ เป็นต้น) รวมถึงเป็นข้อมูลใช้ประกอบการออกแบบแบบฝึกหัดการแปลเพื่อหลีกเลี่ยงรายการภาษาที่เครื่องจักรช่วยแปลสามารถแปลได้ดี

6. ข้อเสนอแนะเพื่อใช้ในการเรียนการสอนและการวิจัย

ผลของงานวิจัยนี้อาจารย์ผู้สอนรายวิชาภาษาต่างประเทศและภาษาไทยสามารถนำผลการวิจัยไปเป็นข้อมูลและสื่อการสอนภาษาและฝึกทักษะการแปลภาษาที่เครื่องจักรช่วยแปลยังไม่สามารถทำได้ดีเทียบเท่ามนุษย์ ส่วนผู้เรียนภาษาต่างประเทศสามารถนำจุดแข็งและจุดอ่อนของเครื่องจักรช่วยแปลไปเป็นเครื่องมือในการเลือกใช้เครื่องจักรช่วยแปลประเภทต่าง ๆ ให้เหมาะสมกับงานแปลภาษาในอนาคต

อีกทั้งอาจารย์ผู้สอนสามารถนำไปประยุกต์ใช้ในรายวิชาการแปลภาษาญี่ปุ่น การอ่านภาษาญี่ปุ่นในส่วนของแปลความหมาย ในส่วนการทำวิจัยต่อยอดสามารถทำได้ในประเด็นเปรียบเทียบความสามารถในการแปลระหว่างเครื่องจักรช่วยแปลประเภทอื่นๆ กับมนุษย์ รวมถึงสามารถต่อยอดงานวิจัยในประเด็นศึกษาพัฒนาความสามารถในการแปลภาษาของเครื่องจักรช่วยแปลโดยงานวิจัยนี้สำรวจความสามารถในการแปลช่วงเดือนตุลาคมถึงพฤศจิกายน พ.ศ. 2561 หากมีการเก็บข้อมูลในแต่ละช่วงเวลาจะทำให้เห็นแนวโน้มพัฒนาการของ Google Translate ระบบ NMT ซึ่งอาศัยฐานข้อมูลขนาดใหญ่ในการวิเคราะห์และสังเคราะห์คำแปลเพื่อให้คณาจารย์ นักศึกษา นักวิจัย และผู้ที่เกี่ยวข้องกับการใช้ภาษามีองค์ความรู้ที่เท่าทันกับการพัฒนาของเทคโนโลยี

เอกสารอ้างอิง/References

- ณัฐริรา ทับทิม. (2557). *ภาษาศาสตร์ภาษาญี่ปุ่นเบื้องต้น*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ณัฐริรา ทับทิม. (2559ก). ลักษณะและความหมายคำประสม NV ในภาษาญี่ปุ่นและภาษาไทย. *JSN Journal*, 6(1), 74-94.
- ณัฐริรา ทับทิม. (2559ข). *โครงสร้างภาษาญี่ปุ่น*. พิษณุโลก: มหาวิทยาลัยราชภัฏพิบูลสงคราม.
- ณัฐริรา ทับทิม. (2560). ลักษณะและความหมายคำประสม VN ในภาษาญี่ปุ่นและภาษาไทย. *JSN Journal*, 7(S), 18-37.
- ณัฐริรา ทับทิม. (2561). *การปรากฏร่วมกันของคำในภาษาญี่ปุ่น: กรณีศึกษาคำนามจากคลังข้อมูลภาษา* (รายงานการวิจัย). พิษณุโลก: มหาวิทยาลัยราชภัฏพิบูลสงคราม.
- ณัฐริรา ทับทิม. (2562). *การปรากฏร่วมกันของคำในภาษาญี่ปุ่น: กรณีศึกษาคำนามจากคลังข้อมูลภาษา*. *JSN Journal*, 9(1), 49-68.
- สมเกียรติ เศวตกิจวณิช. (2549). *แปล ญี่ปุ่น-ไทย (เบื้องต้น)*. กรุงเทพฯ: ภาษาและวัฒนธรรม.
- สุพรรณิ ปิ่นมณี. (2555). *แปลผิด แปลถูก : คัมภีร์การแปลยุคใหม่*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อศินัย ก่อตระกูล และคณะ. (2541). การประมวลผลภาษามนุษย์ด้วยคอมพิวเตอร์. *มนุษยศาสตร์วิชาการ*, 6(1), 94-105.
- 9 อาชีพที่ตกงานก่อนใครเพราะหุ่นยนต์และ AI เข้ามาแทนที่. (2017, มิถุนายน). สืบค้นเมื่อ 31 ธันวาคม 2560, จาก <http://www.theeleader.com/7-job-lost-ai-comming/>
- Bentivogli, L. et al. (2016). Neural versus phrase-based machine translation quality: a case study. *Conference on Empirical Methods in Natural Language Processing*, November 1-5, 2016, Austin, Texas, USA.
- Gotō Isao. (2018). Kikaihonyakugijutsu no kenkyū to dōkō. *NHK Gijutsu R&D*, 168(3), 14-25.
- Group Jammassy. (2554). *พจนานุกรมรูปประโยคภาษาญี่ปุ่น* (พิมพ์ครั้งที่ 1). กรุงเทพฯ: ภาษาและวัฒนธรรม.
- Kōjien. (2008). Tokyo: Iwanamishoten.
- Komeili, Z., Hendavalan, J. F., & Rahimi, A. (2011). An investigation of the translation problems incurred by English-to-Persian Machine Translations: "Padideh, Pars, and Google Softwares". *Procedia-Social and Behavioral Sciences*, (28), 1079-1082.

- Le, Q. V., & Schuster, M. (2016, September 27). A Neural Network for Machine Translation, at Production Scale. Retrieved January 1, 2018, from <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>
- Pearce, D. (2001). Synonymy in Collocation Extraction. Retrieved August 10, 2018, from https://www.researchgate.net/publication/2371642_Synonymy_in_Collocation_Extraction
- Techsource Team. (2560, 14 เมษายน). เจาะลึกการทำงานของ Neural Machine Translation เทคโนโลยีล่าสุดที่ Google Translate ใช้ โดยเลียนแบบระบบคิดจากสมองมนุษย์. สืบค้นเมื่อ 1 มกราคม 2561, จาก <https://techsauce.co/technology/google-translate-neural-machine-learning/>
- Strafella, E. L., Hayashibe, Y., & Matsumoto, Y., (2012). Gendai Nihongo ni okeru koro-ke-shon: kenshuttsu to bunseki. Dai 1 kai ko-pasu Nihongogaku wa-kushoppu, 53-58.
- Wu, Y. et al. (2017). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, 1-23. In A. Villiger, & R. Dach. (Eds.). *International GNSS Service Technical Report 2016 (IGS Annual Report)*. IGS Central Bureau and University of Bern: Bern Open Publishing.