

## การประเมินความถูกต้องในการแปลคำและประโยคทางเภสัชกรรมจากภาษาไทย เป็นอังกฤษของกูเกิลแปลภาษาในปี พ.ศ. 2564 เทียบกับปี พ.ศ. 2560

นริณี โพธิ์ศรี

วิทยาลัยการสาธารณสุขสิรินธร จังหวัดสุพรรณบุรี คณะสาธารณสุขศาสตร์และสหเวชศาสตร์ สถาบันพระบรมราชชนก

### บทคัดย่อ

**วัตถุประสงค์:** เพื่อประเมินความถูกต้องของกูเกิลแปลภาษา (Google Translate: GT) ในการแปลคำและประโยคด้านเภสัชกรรมจากภาษาไทยเป็นอังกฤษในปี พ.ศ. 2560 และ 2564 **วิธีการ:** ผู้วิจัยรวบรวมคำศัพท์และประโยคที่เกี่ยวข้องกับบทสนทนาในร้านยาจากหนังสือและพจนานุกรมด้านสุขภาพและเภสัชกรรม จากนั้น แปลคำศัพท์และประโยคดังกล่าวผ่าน GT ในปี พ.ศ. 2560 และปี พ.ศ. 2564 หลังจากนั้นประเมินความถูกต้องของการแปลโดยผู้ประเมินซึ่งเป็นเภสัชกร 3 ท่าน การประเมินทำใน 2 ระดับตามโครงสร้างของภาษา คือ ระดับคำและระดับประโยค **ผลการวิจัย:** ค่าเฉลี่ยความยอมรับได้ในการแปลคำศัพท์ทางเภสัชกรรมในปี พ.ศ. 2560 และ 2564 มีค่าเท่ากับร้อยละ 76.32 และ 83.42 ตามลำดับ ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ( $P < 0.001$ ) ในระดับประโยคพบว่าค่า adequacy (ผลลัพธ์จากการแปลมีความหมายเหมือนกับข้อความในภาษาต้นแบบ) และค่า fluency (ผลลัพธ์จากการแปลมีความถูกต้องตามหลักไวยากรณ์) ในปี พ.ศ. 2560 มีค่าเท่ากับ  $3.28 \pm 1.46$  และ  $2.97 \pm 1.57$  ตามลำดับ ในขณะที่ปี พ.ศ. 2564 มีค่าเท่ากับ  $4.55 \pm 0.68$  และ  $4.32 \pm 0.91$  ตามลำดับ ซึ่งมีความแตกต่างจากปี พ.ศ. 2560 อย่างมีนัยสำคัญทางสถิติ ( $P < 0.001$ ) การประเมินความถูกต้องด้วยค่า BLEU (BiLingual Evaluation Understudy) ในปี พ.ศ. 2560 และ 2564 มีค่าเท่ากับ  $0.28 \pm 0.41$  และ  $0.41 \pm 0.41$  ตามลำดับ การประเมินผลกระทบด้านลบจากการแปลคลาดเคลื่อนต่อสุขภาพของผู้ป่วยด้วยค่า severity ในปี พ.ศ. 2560 และ 2564 มีค่าเท่ากับ  $3.85 \pm 1.25$  และ  $4.79 \pm 0.48$  ตามลำดับ ( $P < 0.001$ ) (ค่าที่มากหมายถึงมีผลกระทบน้อย) **สรุป:** GT มีการปรับปรุงผลลัพธ์การแปลภาษาให้มีความถูกต้องมากขึ้น อย่างไรก็ตาม การนำเครื่องมือนี้มาใช้ในทางเภสัชกรรมยังจำเป็นต้องตรวจสอบผลลัพธ์ที่ได้จากการแปลว่ามีความถูกต้องเพื่อให้เกิดความมั่นใจว่าจะไม่ส่งผลกระทบต่อสุขภาพของผู้ป่วย

**คำสำคัญ:** กูเกิลแปลภาษา ร้านยา เภสัชกรรม ความถูกต้องในการแปล

รับต้นฉบับ: 20 ต.ค. 2565, ได้รับบทความฉบับปรับปรุง: 29 ธ.ค. 2565, รับลงตีพิมพ์: 30 ธ.ค. 2565

ผู้ประสานงานบทความ: นริณี โพธิ์ศรี วิทยาลัยการสาธารณสุขสิรินธรจังหวัดสุพรรณบุรี คณะสาธารณสุขศาสตร์และสหเวชศาสตร์ สถาบันพระบรมราชชนก E-mail: narinee.pho@phcsuphan.ac.th

## Evaluation of Accuracy in Translating Pharmacy Related Terms and Sentences from Thai to English of the Google Translate in 2021 Compared to That in 2017

Narinee Phosri

Sirindhorn College of Public Health Suphanburi, Faculty of Public Health  
and Allied Health Sciences, Praboromarajchanok Institute

### Abstract

**Objective:** To evaluate the accuracy of Google Translate (GT) in translating terms and sentences related to pharmacy from Thai to English in 2017 and 2021. **Methods:** The researchers collected terms and sentences related to conversations encountered in community pharmacies from books and dictionaries in health and pharmacy. Subsequently, the study translated the terms and sentences through GT in 2017 and 2021. Appraisal of translation accuracy was done by three pharmacists with two levels of language structure i.e., word and sentence levels. **Results:** The average acceptances of translating pharmacy related terms in 2017 and 2021 were 76.32% and 83.42% respectively, with statistically significant differences between years ( $P < 0.001$ ). At the sentence-level, the adequacy and fluency in 2017 were  $3.28 \pm 1.46$  and  $2.97 \pm 1.57$ , respectively, while those in 2021 were  $4.55 \pm 0.68$  and  $4.32 \pm 0.91$ , respectively, with significant differences from those in 2017 ( $P < 0.001$ ). The accuracy evaluation based on the BLEU (BiLingual Evaluation Understudy) in 2017 and 2021 were  $0.28 \pm 0.41$  and  $0.41 \pm 0.41$ , respectively. Assessment of a negative impact of wrong translation on patient health using a severity index revealed that the impacts in 2017 and 2021 was  $3.85 \pm 1.25$  and  $4.79 \pm 0.48$ , respectively ( $P < 0.001$ ) (higher values mean less impact). **Conclusion:** The GT has continuously improved its translation accuracy. However, application of this tool in pharmacy needs verification of translation output to ensure the accuracy and prevent negative impacts on patient health.

**Keywords:** Google Translate, drugstores, pharmacy, translation accuracy

## บทนำ

Machine translation (MT) เป็นการใช้โปรแกรมคอมพิวเตอร์แปลข้อความจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่ง MT เริ่มมีการให้บริการทางเว็บไซต์ในปี ค.ศ. 1996 โดยบริษัท Systran ซึ่งใช้หลักการ rule-based MT คือเป็นการแปลตามหลักภาษาศาสตร์โดยตรงให้มีความถูกต้องตามหลักไวยากรณ์ ต่อมาได้มีการพัฒนาการแปลภาษาของ MT โดยใช้หลักสถิติ (statistical MT) ซึ่งวิเคราะห์ข้อมูลการแปลภาษาของมนุษย์ สร้างคลังข้อมูลเก็บรวบรวมคำศัพท์และรูปแบบการแปลไว้ในฐานข้อมูล และสร้างเป็นแบบจำลองในการแปลภาษา (1) ปัจจุบัน MT ที่ให้บริการในรูปแบบเชิงพาณิชย์มีหลากหลาย เช่น Systran translate, Amazon translate และ memoQ translator และให้บริการโดยไม่คิดค่าใช้จ่าย เช่น กูเกิลแปลภาษา (Google Translate: GT), Bing Microsoft translator และ DeepL translator เป็นต้น

ในทางการแพทย์การสื่อสารด้วยภาษาที่แตกต่างกันถือเป็นอุปสรรคที่สำคัญอย่างหนึ่งในการบริการผู้ป่วย (2-5) GT เป็นโปรแกรมแปลภาษาออนไลน์ที่ได้รับความนิยมอย่างแพร่หลาย ถือเป็นเครื่องมือหนึ่งที่ใช้ในการแก้ปัญหาดังกล่าว เนื่องจากสามารถใช้บริการโดยไม่เสียค่าใช้จ่าย ใช้งานง่าย สะดวก แปลได้หลากหลายภาษา และสามารถช่วยลดระยะเวลาในการแปลภาษาได้ (6-8)

หลายการศึกษาพบว่า GT ยังมีความถูกต้องไม่เพียงพอในการแปลภาษาในทางสุขภาพ (3, 9-12) ในปี พ.ศ. 2559 ผู้วิจัยศึกษาความถูกต้องของ GT ในการแปลภาษาด้านเภสัชกรรมจากภาษาไทยไปเป็นอังกฤษพบว่า GT มีความสามารถในการการแปลคำศัพท์ถูกต้องเพียงร้อยละ 53 (13) อย่างไรก็ตาม การศึกษานี้รวบรวมคำศัพท์จากหนังสือเภสัชศาสตร์ อังกฤษ-ไทย ไทย-อังกฤษ ฉบับราชบัณฑิตยสถานในปี พ.ศ. 2551 (14) เพียงแหล่งเดียว ทำให้มีจำนวนคำศัพท์น้อยเพียง 446 คำ ในปี พ.ศ. 2560 ผู้วิจัยได้ประเมินความถูกต้องของ GT ในการแปลข้อความที่ใช้สื่อสารในร้านยาระหว่างภาษาไทย อังกฤษ และพม่า พบว่า GT ยังไม่มีประสิทธิภาพในการแปลภาษาด้านเภสัชกรรมจากภาษาไทยไปเป็นภาษาอังกฤษ ซึ่งพบทั้งในระดับคำศัพท์และระดับประโยค (12, 15)

ในเดือนพฤศจิกายน ปี พ.ศ. 2559 GT ได้ปรับเปลี่ยนแบบจำลองในการแปลภาษาจากหลักสถิติเป็นแบบจำลองการแปลภาษาโดยใช้ปัญญาประดิษฐ์ (artificial

intelligence: AI) ที่เรียกว่า neural machine translation (NMT) การแปลจะพิจารณาประโยคทั้งประโยคแทนที่จะตัดเป็นคำหรือวลีมาแปลทีละส่วน ทำให้การแปลภาษามีความใกล้เคียงกับภาษาของมนุษย์และมีความเป็นธรรมชาติมากที่สุด (8) ปัจจุบันด้วยความก้าวหน้าของเครือข่ายอินเทอร์เน็ต โปรแกรมสามารถเก็บข้อมูลการแปลในระบบออนไลน์ ซึ่งมีความสะดวก รวดเร็ว และสามารถเก็บข้อมูลได้เป็นจำนวนมาก ฉะนั้นเมื่อเวลาผ่านไป GT มีคลังข้อมูลของคำศัพท์ที่ถูกแปลและมีข้อมูลการแปลที่หลากหลายรูปแบบมากขึ้น โดยในปี พ.ศ. 2560 GT ให้บริการแปลภาษาทั้งหมด 103 ภาษา และเพิ่มเป็น 109 ภาษาในปี พ.ศ. 2564 GT ตลอดจนมีการปรับปรุงค่าแปลจากการเก็บรวบรวมข้อมูลการแปลที่มากขึ้น

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอความถูกต้องของ GT ในการแปลภาษาด้านเภสัชกรรมจากไทยเป็นอังกฤษในปี พ.ศ. 2560 ซึ่งเป็นช่วงเริ่มแรกในการใช้หลักการ NMT และในปี พ.ศ. 2564 ที่ GT มีการเก็บข้อมูลการแปลที่เพิ่มขึ้น เพื่อให้มั่นใจได้ว่า ผลลัพธ์จากการแปลภาษามีความถูกต้องหากนำมาใช้ในทางเภสัชกรรม

## วิธีการวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงพรรณนาซึ่งได้รับการยกเว้นการพิจารณาจริยธรรมการวิจัยในมนุษย์จากวิทยาลัยการสาธารณสุขสิรินธร จังหวัดสุพรรณบุรี คณะสาธารณสุขศาสตร์และสหเวชศาสตร์ สถาบันพระบรมราชชนก เลขที่ PHCSP-A2563/01-14

### แหล่งที่มาของคำศัพท์และประโยค

ผู้วิจัยรวบรวมคำศัพท์และประโยคที่เกี่ยวข้องกับบทสนทนาในร้านยาจากหนังสือและพจนานุกรมด้านสุขภาพและเภสัชกรรม (14, 16-20) คำทับศัพท์จะไม่ถูกคัดเลือกมาประเมิน

### การแปลภาษาและการประเมินผล

คำและประโยคที่รวบรวมได้ถูกแปลจากภาษาไทยเป็นภาษาอังกฤษผ่าน GT ในช่วงปี พ.ศ. 2560 คำและประโยคเหล่านี้ถูกนำมาแปลผ่าน GT อีกครั้งในช่วงปี พ.ศ. 2564 การประเมินผลการแปลใน 2 ช่วงทำโดยเภสัชกรสามท่านที่จบการศึกษาจากต่างประเทศในระดับปริญญาโทหรือปริญญาเอก ซึ่งมีความสามารถในการใช้ภาษาอังกฤษเป็นอย่างดี การประเมินแบ่งเป็น 2 ระดับ ได้แก่ ระดับคำและระดับประโยค คำ “ความยอมรับได้” ใช้ประเมินความถูกต้อง

ของการแปลในระดับคำ คำนี้หมายถึงการที่คำแปลสามารถเข้าใจหรือนำไปใช้ในทางเภสัชกรรมได้ ผลการประเมินมี 2 แบบ คือ ยอมรับได้หรือไม่สามารถยอมรับได้ ค่าความยอมรับได้ตั้งแต่ร้อยละ 80 ขึ้นไปแสดงว่า ผลการแปลภาษาถูกต้องในระดับที่ดีมาก (21)

ค่า adequacy และค่า fluency เป็นการประเมินความถูกต้องในระดับประโยค โดยค่า adequacy หมายถึงผลลัพธ์จากการแปลมีความหมายเหมือนกับข้อความในภาษาต้นแบบ (9) ค่า fluency หมายถึง ผลลัพธ์จากการแปลมีความถูกต้องตามหลักไวยากรณ์ (9) นอกจากนี้ ผู้ประเมินยังได้ประเมินค่า severity ซึ่งแสดงถึง ผลกระทบด้านลบต่อสุขภาพของผู้ป่วยอันเนื่องมาจากผลลัพธ์ของการแปล (9) ค่า adequacy ค่า fluency และค่า severity มีผลการประเมิน 5 ระดับ ความหมายของระดับคะแนนต่าง ๆ แสดงในตารางที่ 1

ค่า BiLingual Evaluation Understudy (BLEU) เป็นการหาค่าความถูกต้องตามหลักการทางภาษาศาสตร์ที่เรียกว่า N-gram หรือจำนวนคำ เป็นการวัดประสิทธิภาพในการแปลซึ่งมีความใกล้เคียงกับการแปลโดยมนุษย์ การคำนวณค่า BLEU ทำโดยนำผลการแปลจาก GT มาเปรียบเทียบกับประโยคอ้างอิงในหนังสือ และพิจารณาจำนวนคำที่มีผลการแปลตรงกันตั้งแต่ 1 คำ จนถึง 4 คำ จากนั้นนำมาคำนวณโดยใช้โปรแกรมอัตโนมัติ (22)

การประเมินประสิทธิภาพการแปลภาษาใช้ผู้ประเมิน 3 คนโดยประเมินแบบอิสระต่อกัน ผู้วิจัยประเมินความสอดคล้องกันของผู้ประเมินโดยใช้สถิติ Fleiss's Kappa (23) ค่าคะแนน <0, 0.01 – 0.20, 0.21 – 0.40, 0.41 – 0.60, 0.61 – 0.80, และ 0.80 – 1.00 หมายถึง ไม่มี ความสอดคล้อง สอดคล้องเล็กน้อย พอใช้ ปานกลาง ดี และดีมากตามลำดับ

**ตารางที่ 1.** ความหมายของค่า adequacy ค่า fluency และค่า severity

ระดับ	adequacy	fluency	severity
1	ไม่เข้าใจ	ไม่มีความถูกต้องทางไวยากรณ์	ส่งผลอันตรายต่อผู้ป่วย
2	เข้าใจน้อย	มีความถูกต้องทางไวยากรณ์น้อย	ทำให้ผู้ป่วยไม่ได้รับการดูแลรักษาที่จำเป็น
3	เข้าใจค่อนข้างมาก	มีความถูกต้องทางไวยากรณ์ค่อนข้างมาก	ทำให้การดูแลรักษาล่าช้า
4	เข้าใจมาก	มีความถูกต้องทางไวยากรณ์มาก	ส่งผลกระทบต่อผู้ป่วยแบบไม่ชัดเจน
5	เข้าใจอย่างสมบูรณ์	มีความถูกต้องทางไวยากรณ์ทั้งหมด	ไม่มีผลกระทบต่อผู้ป่วย

**การวิเคราะห์ผล**

ร้อยละของการเปลี่ยนแปลงผลลัพธ์ของการแปลแสดงถึงจำนวนคำศัพท์ภาษาอังกฤษซึ่งเป็นผลลัพธ์จากการแปลคำศัพท์ภาษาไทยผ่าน GT ในปี พ.ศ. 2564 ที่แตกต่างไปจากปีพ.ศ. 2560 ผลการประเมินจากผู้ประเมินทั้ง 3 คนด้วยค่าความยอมรับได้ในระดับคำ ค่า adequacy ค่า fluency และค่า severity ในระดับประโยคจะถูกนำเสนอผลด้วยค่าเฉลี่ย เนื่องจากการกระจายของข้อมูลเป็นแบบไม่ปกติ จึงใช้สถิติทดสอบ Wilcoxon signed rank test ในการเปรียบเทียบผลความถูกต้องของการแปลระหว่างปี พ.ศ. 2560 กับ 2564

**ผลการวิจัย**

คำศัพท์ด้านเภสัชกรรมที่รวบรวมได้มีทั้งหมด 870 คำ แบ่งเป็น 6 หมวดหมู่ ได้แก่ อวัยวะและสารต่าง ๆ ในร่างกาย 125 คำ โรคและอาการ 348 คำ ยาและผลิตภัณฑ์สุขภาพ 130 คำ การบริหารยา 86 คำ รูปแบบเภสัชภัณฑ์ 98 คำ และเครื่องมือแพทย์และเครื่องสำอาง 83 คำ ประโยคด้านเภสัชกรรมมีจำนวนทั้งหมด 100 ประโยค แบ่งตามประเภทของการสนทนาออกเป็น 3 ประเภท ได้แก่ การซักประวัติผู้ป่วย 40 ประโยค การบริหารยา 40 ประโยค และการให้คำแนะนำในการปฏิบัติตัว 20 ประโยค

**ผลการประเมินในระดับคำ**

จากตารางที่ 2 ผลลัพธ์การแปลภาษาของ GT ใน ปี พ.ศ. 2564 มีการเปลี่ยนแปลงค่าแปลจากปี พ.ศ. 2560 จำนวน 279 คำ จากจำนวนคำ 870 คำ คิดเป็นร้อยละ 32.07 คำศัพท์ประเภทเครื่องมือแพทย์และเครื่องสำอางมีการเปลี่ยนแปลงในการแปลมากที่สุด จำนวน 42 คำจากจำนวนคำ 83 คำ คิดเป็น ร้อยละ 50.60 รองต่อมา คือ คำ

**ตารางที่ 2.** ผลการประเมินในระดับคำด้วยค่าความยอมรับได้

ประเภทของคำศัพท์ด้านเภสัชกรรม (จำนวนคำ)	จำนวนคำในปี พ.ศ. 2564 ที่มีค่าแปลต่างจาก ปี พ.ศ. 2560	ร้อยละของคำที่มี การเปลี่ยนแปลง ค่าแปล	ร้อยละของค่าเฉลี่ย ความยอมรับได้		P <sup>1</sup>
			2560	2564	
อวัยวะและสารต่าง ๆ ในร่างกาย (125)	10	8.00	90.67	93.33	0.03
โรคและอาการ (348)	121	34.77	79.15	82.23	0.12
ยาและผลิตภัณฑ์สุขภาพ (130)	63	48.46	60.51	74.87	< 0.001
การบริหารยา (86)	12	13.95	86.82	89.53	0.30
รูปแบบเภสัชภัณฑ์ (98)	31	31.63	72.11	80.95	< 0.001
เครื่องมือแพทย์และเครื่องสำอาง (83)	42	50.60	61.90	83.33	< 0.001
รวมทั้งหมด (870)	279	32.07	76.32	83.42	< 0.001

1: Wilcoxon signed rank test

ศัพท์ที่เกี่ยวข้องกับยาและผลิตภัณฑ์สุขภาพมีการเปลี่ยนแปลงจำนวน 63 คำจากจำนวนคำ 130 คำ (ร้อยละ 48.46)

ค่าเฉลี่ยความยอมรับได้ในการแปลคำศัพท์ในปี พ.ศ. 2560 และปี พ.ศ. 2564 มีค่าเท่ากับ 76.32 และ 83.42 ตามลำดับ และมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ( $P < 0.001$ ) เมื่อพิจารณาตามหมวดหมู่ คำศัพท์ทุกประเภทการแปลใน 2 ช่วงเวลา มีความถูกต้องของการแปลแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ยกเว้นคำศัพท์ในหมวดหมู่โรคและอาการ และหมวดหมู่การบริหารยา ร้อยละของค่าเฉลี่ยความยอมรับได้ในปี พ.ศ. 2560 มีเพียงคำศัพท์ที่เกี่ยวข้องกับอวัยวะและสารต่าง ๆ ในร่างกายและการบริหารยาที่มีค่ามากกว่าร้อยละ 80 โดยมีค่าเฉลี่ยความยอมรับได้เท่ากับร้อยละ 90.67 และร้อยละ 86.82 ตามลำดับ คำศัพท์เกี่ยวกับยาและผลิตภัณฑ์สุขภาพมีค่าเฉลี่ยความยอมรับได้น้อยที่สุดเท่ากับร้อยละ 60.51 รองต่อมา คือ คำศัพท์ที่เกี่ยวข้องกับเครื่องมือแพทย์และเครื่องสำอาง (ร้อยละ 61.90)

ในปี พ.ศ. 2564 คำศัพท์ทุกหมวดหมู่มีร้อยละของค่าเฉลี่ยความยอมรับได้มากกว่าร้อยละ 80 ยกเว้นคำศัพท์เกี่ยวกับยาและผลิตภัณฑ์สุขภาพซึ่งมีค่าเฉลี่ยความยอมรับได้น้อยที่สุดเท่ากับร้อยละ 74.87 คำศัพท์เกี่ยวกับอวัยวะและสารต่าง ๆ ในร่างกายมีความถูกต้องมากที่สุด (ร้อยละ 93.33) ค่าสถิติ Fleiss's Kappa ซึ่งวัดความสอดคล้องของความเห็นจากผู้ประเมิน มีค่าเท่ากับ 0.71 แสดงถึงการประเมินมีความสอดคล้องในระดับดี

**ผลการประเมินในระดับประโยค**

ในปี พ.ศ. 2560 ค่า adequacy และค่า fluency ของการแปลประโยครวมทั้งหมดมีค่าเท่ากับ  $3.28 \pm 1.46$  และ  $2.97 \pm 1.57$  ตามลำดับ (ตารางที่ 3) ในขณะที่ในปี พ.ศ. 2564 มีค่าเท่ากับ  $4.55 \pm 0.68$  และ  $4.32 \pm 0.91$  ตามลำดับ ซึ่งความถูกต้องมีความแตกต่างจากปี พ.ศ. 2560 อย่างมีนัยสำคัญทางสถิติ ( $P < 0.001$ )

เมื่อพิจารณาในแต่ละประเภทของประโยค ค่า adequacy และค่า fluency ในปี พ.ศ. 2564 มากกว่าปี พ.ศ. 2560 ในทุกประเภทของประโยคอย่างมีนัยสำคัญทางสถิติ โดยประโยคการชักประวัติผู้ป่วยมีความถูกต้องน้อยที่สุดในขณะที่ประโยคการให้คำแนะนำในการปฏิบัติตัวมีความถูกต้องมากที่สุดทั้งในปี พ.ศ. 2560 และปี พ.ศ. 2564

การประเมินผลกระทบด้านลบต่อสุขภาพของผู้ป่วยจากการแปลผิดด้วยค่า severity ในปี พ.ศ. 2560 มีค่าเท่ากับ  $3.85 \pm 1.25$  หมายถึงการแปลทำให้การดูแลรักษาล่าช้าและปี พ.ศ. 2564 มีค่าเท่ากับ  $4.79 \pm 0.48$  หมายถึงการแปลส่งผลกระทบต่อผู้ป่วยแบบไม่ชัดเจน โดยมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ( $P < 0.001$ ) ความสอดคล้องกันของผู้ประเมินจากค่าสถิติ Fleiss's Kappa ในมีค่าเท่ากับ 0.36 แสดงถึงการประเมินมีความสอดคล้องในระดับพอใช้

ค่า BLEU ของการแปลประโยคทั้งหมด 100 ประโยคในปี พ.ศ. 2564 มีการเปลี่ยนแปลงจากปี พ.ศ. 2560 อย่างมีนัยสำคัญทางสถิติ ( $P = 0.01$ ) โดยค่า BLEU

**ตารางที่ 3.** ผลการประเมินในระดับประโยคด้วยค่า adequacy ค่า fluency และค่า severity

	การชักประวัติผู้ป่วย (N = 40)	การให้คำแนะนำในการใช้ยา (N = 40)	การให้คำแนะนำในการปฏิบัติตัว (N = 20)	รวมทั้งหมด (N = 100)
<b>adequacy</b>				
2560	2.31±1.42	3.87±1.9	4.05±1.08	3.28±1.46
2564	4.45±0.87	4.63±0.50	4.62±0.54	4.55±0.68
P <sup>1</sup>	< 0.001	< 0.001	0.012	< 0.001
<b>fluency</b>				
2560	1.75±1.24	3.63±1.34	3.77±1.14	2.97±1.57
2564	4.06±1.12	4.56±0.66	4.38±0.79	4.32±0.91
P <sup>1</sup>	< 0.001	< 0.001	0.010	< 0.001
<b>severity</b>				
2560	3.03±1.36	4.38±0.83	4.42±0.74	3.85±1.25
2564	4.79±0.53	4.79±0.51	4.78±0.27	4.79±0.48
P <sup>1</sup>	< 0.001	< 0.001	0.045	< 0.001

1: Wilcoxon signed rank test ที่เปรียบเทียบค่าในปี พ.ศ. 2560 กับปี พ.ศ. 2564

ในปี พ.ศ. 2560 และ 2564 มีค่าเท่ากับ 0.28±0.41 และ 0.41±0.41 ตามลำดับ โดยประโยคที่เกี่ยวกับการชักประวัติผู้ป่วยมีค่า BLEU ในปี พ.ศ. 2564 มากกว่าในปี พ.ศ. 2560 แต่ประโยคประเภทอื่น มีค่า BLEU ในปี พ.ศ. 2564 และ 2560 ที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

### การอภิปรายผล

ความถูกต้องในการแปลภาษาทางเภสัชกรรมจากภาษาไทยไปเป็นภาษาอังกฤษของ GT ในปี พ.ศ. 2564 มีการเปลี่ยนแปลงจากปี พ.ศ. 2560 โดยสอดคล้องกับการศึกษาของ Milam ที่ประเมินความถูกต้องในการแปลของ GT ในปี ค.ศ. 2011 ทั้งหมด 51 ภาษา พบว่า การแปลภาษาในแถบยุโรปมีความถูกต้องค่อนข้างมาก แต่

ภาษาในแถบเอเชียยังไม่ถูกต้องมากนัก แต่เมื่อทำการประเมินอีกครั้งในปี ค.ศ. 2019 พบว่า ความถูกต้องเพิ่มมากขึ้นถึงร้อยละ 34 (24)

สิ่งที่สำคัญในการใช้ GT คือ ควรหลีกเลี่ยงการแปลคำศัพท์ในหมวดหมู่ยาและผลิตภัณฑ์สุขภาพ เนื่องจากคำศัพท์ในประเภทนี้ส่วนใหญ่เป็นคำศัพท์เฉพาะทาง ส่งผลให้การแปลไม่ถูกต้องเท่าที่ควร ในปี พ.ศ. 2564 ค่าเฉลี่ยความยอมรับเท่ากับ 74.87 ซึ่งน้อยกว่าร้อยละ 80 อย่างไรก็ตามการศึกษานี้เก็บข้อมูลผลลัพธ์จากการแปลของ GT โดยเลือกที่ผลลัพธ์หลักที่โปรแกรมแสดง ไม่ได้เก็บข้อมูลผลลัพธ์ย่อยที่แสดงข้อมูลอยู่ด้านล่างของผลลัพธ์หลัก ซึ่งจะเรียงตามลำดับความถี่ของคำแปลที่ปรากฏในเอกสารสาธารณะจากมากไปน้อย ดังนั้นอาจมีคำแปลที่สามารถ

**ตารางที่ 4.** ผลการประเมินในระดับประโยคด้วยค่า BLEU

ประเภทของประโยค (จำนวนประโยค)	BLEU		P <sup>1</sup>
	ปี พ.ศ. 2560	ปี พ.ศ. 2564	
การชักประวัติผู้ป่วย (40)	0.19±0.34	0.49±0.41	< 0.001
การบริหารยา (40)	0.58±0.29	0.66±0.38	0.42
การให้คำแนะนำในการปฏิบัติตัว (20)	0.45±0.54	0.39±0.49	0.51
รวมทั้งหมด (100)	0.28±0.41	0.41±0.41	0.01

1: Wilcoxon signed rank test

นำมาใช้ทางเภสัชกรรมได้ หากมีการเก็บข้อมูลผลลัพธ์การแปลทั้งจากผลลัพธ์หลักและผลลัพธ์ย่อย น่าจะมีร้อยละของค่าเฉลี่ยความยอมรับได้มากกว่าที่รายงานในผลการศึกษา

จากผลการศึกษาแสดงให้เห็นว่าในปี พ.ศ. 2564 GT มีการเปลี่ยนแปลงผลลัพธ์ในการแปลในระดับประโยค โดยเฉพาะประโยคคำถาม เนื่องจากค่า adequacy ค่า fluency และค่า BLEU ของประโยคที่เกี่ยวกับการซักประวัติ ในปี พ.ศ. 2564 มากกว่าปี พ.ศ. 2560 อย่างมีนัยสำคัญทางสถิติ ซึ่งประโยคในการซักประวัติผู้ป่วยมีโครงสร้างประโยคเป็นประโยคคำถาม ผลลัพธ์ในการแปลประโยคที่มีโครงสร้างเป็นประโยคคำถามในปี พ.ศ. 2560 ของ GT ส่วนใหญ่จะแสดงผลการแปลเป็นประโยคบอกเล่า ทำให้เกิดการแปลผิดโครงสร้างประโยคและการสื่อความหมายที่ผิดเพี้ยน ตัวอย่างเช่น “คุณรู้สึกเวียนศีรษะไหม” มีคำแปลในปี พ.ศ. 2560 คือ “You feel dizzy silk” ส่วนในปี พ.ศ. 2564 GT ได้มีการปรับปรุงผลลัพธ์เป็น “Do you feel dizzy” ซึ่งมีความถูกต้องทั้งความหมายและไวยากรณ์

การเริ่มต้นสำรวจและเก็บคำศัพท์ในคลังข้อมูลตามหลักการของการแปลภาษาโดยเครื่องจักรอัตโนมัติจะใช้หลักการการเก็บข้อมูลการแปลภาษาของมนุษย์ที่ทำการแปลภาษาต่าง ๆ จากทั่วโลก คำศัพท์ที่ใช้สร้างแบบจำลองในการแปลภาษาเริ่มจากภาษาอังกฤษเพราะภาษาอังกฤษเป็นภาษากลางในการสื่อสารของคนทั่วโลก ปัจจุบันโปรแกรมแปลภาษาสามารถเก็บข้อมูลการแปลผ่านระบบเครือข่ายอินเทอร์เน็ต ซึ่งทำให้เก็บข้อมูลการแปลเข้าไปวิเคราะห์รูปแบบการแปลภาษาเป็นจำนวนมากได้ ฉะนั้นภาษาที่ถูกใช้มากและมีข้อมูลการแปลในอินเทอร์เน็ตมากจะถูกนำมาสร้างแบบจำลองซึ่งมีความสลับซับซ้อนได้มากกว่าภาษาที่มีข้อมูลการแปลน้อย ข้อมูลจาก Internet World Stats ปี พ.ศ. 2564 พบว่า ภาษาอังกฤษเป็นภาษาที่ใช้มากที่สุดบนอินเทอร์เน็ต (25) จึงทำให้ผลลัพธ์ในการแปลมีการเปลี่ยนแปลงแก้ไขจากการเก็บสถิติที่เพิ่มขึ้นในคลังข้อมูล และส่งผลให้การแปลมีความถูกต้องมากขึ้น ตัวอย่างเช่น ยาลดไข้ ในปี พ.ศ. 2560 GT แปลเป็นภาษาอังกฤษเป็นคำว่า “aspirin” แต่ในปี พ.ศ. 2564 แปลได้ผลลัพธ์เป็น “antipyretics”

ค่าสถิติ Fleiss's Kappa แสดงให้เห็นได้ว่า ในระดับคำ ผลการประเมินจากผู้ประเมินมีความสอดคล้องกันอยู่ในระดับดี แต่ในระดับประโยคกับพบว่าอยู่ในระดับพอใช้ ที่

เป็นเช่นนี้เนื่องจากการประเมินในระดับประโยคนั้น ผู้ประเมินสามารถให้ค่าคะแนนได้ถึง 5 ระดับ ส่วนในระดับคำ การให้ค่าคะแนนมีเพียงแค่ 2 ระดับ จึงมีโอกาสที่การให้ค่าคะแนนจะไม่ตรงกันได้

อย่างไรก็ตามหากมีความจำเป็นต้องใช้ GT ในทางเภสัชกรรม ควรใช้ GT เป็นเครื่องมือเบื้องต้นเพื่อช่วยลดระยะเวลาในการแปล และควรมีการตรวจสอบความถูกต้องก่อนนำคำแปลที่ได้มาใช้จริง (3) Turner และคณะแนะนำให้ใช้ GT ในขั้นตอนแรกเพื่อช่วยลดระยะเวลาในการแปลและช่วยประหยัดค่าใช้จ่ายในการแปล หลังจากนั้นต้องมีการแก้ไขคำแปลจากมนุษย์ให้ถูกต้อง (3) Taira และคณะ กล่าวว่า ถึงแม้จะพบว่า การแปลจาก GT ให้ค่า adequacy และ fluency สูงแต่ก็ไม่ควรนำมาใช้ในการแปลทางคลินิกเนื่องจากความถูกต้องไม่เท่ากันในแต่ละภาษา (26)

### สรุปผลและข้อเสนอแนะ

GT มีการพัฒนาและปรับปรุงผลลัพธ์การแปลภาษาทางเภสัชกรรมจากการแปลภาษาไทยไปเป็นภาษาอังกฤษให้มีความถูกต้องมากขึ้น อย่างไรก็ตามการนำ GT มาใช้ จำเป็นต้องมีการปรับปรุงแก้ไขผลลัพธ์ที่ได้จากการแปลของเครื่องมือนี้ก่อนนำมาประยุกต์ใช้สื่อสารทางเภสัชกรรม งานวิจัยต่อไปควรเพิ่มจำนวนคำและประโยคให้ครอบคลุมเนื้อหาทางด้านเภสัชกรรมทั้งหมด หรือประเมินความถูกต้องในระดับที่สูงขึ้น เช่น ระดับเอกสารหรือระดับบทความ เป็นต้น นอกจากนี้ สำหรับการวิจัยในอนาคต ควรตรวจสอบผลลัพธ์ของการแปลภาษาระดับประโยคด้วย GT โดยให้ตัวอย่างชาวต่างชาติที่ใช้ภาษาอังกฤษในชีวิตประจำวันอธิบายความหมายคำแปลดังกล่าวตามความเข้าใจของตน เพื่อให้ผู้วิจัยตัดสินใจว่ามีความหมายตรงกับภาษาไทยหรือไม่

### เอกสารอ้างอิง

1. Okpor M. Machine translation approaches: issues and challenges. Int J Comput Sci Appl 2014; 11: 159-65.
2. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a

- systematic review. *Oman Med J* 2020; 35: e122. doi: 10.5001/omj.2020.40.
3. Dew KN, Turner AM, Choi YK, Bosold A, Kirchoff K. Development of machine translation technology for assisting health communication: A systematic review. *J Biomed Inform* 2018; 85: 56-67.
  4. Li C, Son N, Abdulkerim B, Jordan CA, Son CGE. Overcoming communication barriers to healthcare for culturally and linguistically diverse patients. *N Am J Med Sci* 2017; 10: 103-9.
  5. Mohammad A, Saini B, Chaar BB. Pharmacists' experiences serving culturally and linguistically diverse patients in the Australian community pharmacy setting. *Int J Clin Pharm* 2021; 43: 1563-73.
  6. Google. Google Translate [online]. 2017 [cited Aug 10, 2017]; Available from: [translate.google.com?hlth](https://translate.google.com?hlth).
  7. Rodriguez JA, Fossa A, Mishuris R, Herrick B. Bridging the language gap in patient portals: An evaluation of Google translate. *J Gen Intern Med* 2021; 36: 567-9.
  8. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. Google's neural machine translation system: Bridging the gap between human and machine translation [online]. 2016 [cited May 2, 2017]. Available from: [arxiv.org/abs/1609.08144](https://arxiv.org/abs/1609.08144).
  9. Chen X, Acosta S, Barry AE. Evaluating the accuracy of Google Translate for diabetes education material. *JMIR diabetes* 2016; 1: e3. doi:10.2196/diabetes.5848.
  10. Cornelison BR, Al-Mohaish S, Sun Y, Edwards CJ. Accuracy of Google Translate in translating the directions and counseling points for top-selling drugs from English to Arabic, Chinese, and Spanish. *Am J Health Syst Pharm* 2021; 78: 2053-8.
  11. Patil S, Davies P. Use of Google translate in medical communication: evaluation of accuracy. *BMJ* 2014; 349: g7392. doi: 10.1136/bmj.g7392
  12. Phosri N, Lertnattee V. Google translate using for a communication tool between Thai pharmacists and Myanmar patients in drugstores. *J Hlth Sci Res* 2019.13: 104-14.
  13. Phosri N, Lertnattee V. An evaluation of Google Translate in pharmaceutical language: Thai and English translations. The 6th international graduate study conference: Creative multi-disciplinary studies for sustainable development; 2016 July 11- 12; Bangkok,Thailand. Bangkok: Silpakorn University; 2016. p. S1009 – 19.
  314. Royal Institute. English-Thai, Thai-English pharmaceutical vocabulary Royal institute dictionary. 2nd ed. Bangkok: Aksorncharoentat; 2008.
  15. Phosri N, Lertnattee V. An evaluation of Google Translate in pharmacy domain between Thai and Myanmar. *Adv Sci Lett* 2018; 24: 2746-50.
  16. Nahler G. Dictionary of pharmaceutical medicine. Vienna: Springer; 2009.
  17. Bureau of Health Administration. Manual in four languages Thai-English-Myanmar-Cambodian for health personnel in providing services to foreigners. Nonthaburi: Bureau of Health Administration; 2014.
  18. Chanyam W. Handbook of medical conversation; English conversation for nurses; Conversation etc. without fear ; Traveling conversation for army medical personnel. Bangkok: Army Medical Field Service School; 2013.
  19. Royal Institute. English-Thai, the pharmaceutical dictionary. 2 nd ed. Bangkok: Royal Institute; 2008.
  20. Eiamprapai P. The handbook of ASEAN language for medical personnel. Humanities and Social Sciences Mahasarakham University: Mahasarakham ; 2014.
  21. Costa-jussà MR, Farrús M, Pons JS. Machine translation in medicine. In: Mokryš M,Lieskovský A, editors. Proceedings of the 1st virtual international conference on advanced research in scientific fields; 2012 Dec 3- 7; Žilina, Slovakia. Žilina: Advanced research in scientific areas; 2012. p.1995-8.
  22. Seljan S, Brkić M, Vičić T. BLEU evaluation of machine-translated English-Croatian legislation.

- Proceedings of the eighth international conference on language resources and evaluation; 2012 May; Istanbul, Turkey. Istanbul: European language resources association; 2012.p. 2143-8.
23. Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. *Qual Quant* 2015; 49: 463-70.
24. Aiken M. An updated evaluation of Google translate accuracy. *Studies in Linguistics and Literature*. 2019; 3: 253-60.
25. Internet World Stats. Internet world users by language [online]. 2021 [ cited Dec 25, 2021] ; Available from: [www.Internetworldstats.com/stats7.htm](http://www.Internetworldstats.com/stats7.htm)
26. Taira BR, Kreger V, Orue A, Diamond LC. A pragmatic assessment of Google Translate for emergency department instructions. *J Gen Intern Med*. 2021; 36: 3361-5.