

การประเมินคุณภาพในการแปลภาษาระหว่างภาษาจีนและภาษาไทยสำหรับ การสื่อสารเพื่อชกประวัติเบื้องต้นของภูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีที

ภาพิต มากคุณ¹, วีรยุทธ์ เลิศนที²

¹นักศึกษาปริญญาโท หลักสูตรเภสัชศาสตรมหาบัณฑิต สาขาสารสนเทศศาสตร์ทางสุขภาพ

คณะเภสัชศาสตร์มหาวิทยาลัยศิลปากร

²สาขาสุขภาพดิจิทัล คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

บทคัดย่อ

วัตถุประสงค์: เพื่อประเมินและเปรียบเทียบคุณภาพการแปลภาษาระหว่างภาษาจีนและภาษาไทยสำหรับการชกประวัติเบื้องต้นของภูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีที **วิธีการ:** การวิจัยเชิงประเมินประสิทธิภาพนี้เริ่มจากการทดสอบคำสั่งหรือ prompt สำหรับแชตจีพีที 3 รูปแบบ เพื่อเลือกคำสั่งที่เหมาะสมที่สุดสำหรับการแปลข้อความด้านสุขภาพ จากนั้นจึงใช้ภูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีทีแปลคำศัพท์เกี่ยวกับโรคและอาการจำนวน 75 คำ และวลีหรือประโยคจำนวน 86 วลีหรือประโยคที่เกี่ยวข้องกับการชกประวัติเบื้องต้น ระบบประสาท ระบบทางเดินอาหาร และระบบสืบพันธุ์ ผู้วิจัยได้คัดเลือกคำ วลี หรือประโยคดังกล่าวจากคู่มือการชกประวัติผู้รับบริการ การแปลมีทั้งการแปลภาษาจีนเป็นภาษาไทยและภาษาไทยเป็นภาษาจีน การประเมินคุณภาพการแปลดำเนินการโดยผู้เชี่ยวชาญ 2 ท่าน การประเมินแบ่งเป็น 2 ระดับ ได้แก่ 1) ระดับคำศัพท์ ประเมินด้วยค่ายอมรับได้ และเปรียบเทียบคุณภาพการแปลของเครื่องมือทั้งสามด้วย Bayesian generalized linear mixed models และ 2) ระดับวลีหรือประโยค ประเมินด้วยแบบสอบถามมาตรวัดลิเคิร์ท 5 ระดับ ใน 4 ด้าน ได้แก่ adequacy (ความครบถ้วน), fluency (ความเป็นไปตามไวยากรณ์และเข้าใจได้), meaning (ความหมาย) และ severity (ผลกระทบต่อสุขภาพ) การเปรียบเทียบคุณภาพการแปลของเครื่องมือทั้งสามใช้ ordinal logistic mixed-effects model **ผลการวิจัย:** การแปลระดับคำศัพท์ไม่พบปฏิสัมพันธ์ระหว่างเครื่องมือแปลและผู้ประเมินอย่างมีนัยสำคัญทางสถิติ แชตจีพีทีมีโอกาสได้คำแปลที่ยอมรับได้สูงกว่าเครื่องมืออื่นอย่างมีนัยสำคัญทางสถิติทั้งในทิศทางจีนเป็นไทยและไทยเป็นจีน อัตราการยอมรับการแปลคำศัพท์ได้ของแชตจีพีทีเท่ากับร้อยละ 99.98 และ 99.96 ตามลำดับ แชตจีพีทีที่มีคุณภาพการแปลในระดับคำศัพท์ที่ดีกว่าภูเกิลแปลภาษา และไปตู้แปลภาษาอย่างมีนัยสำคัญทางสถิติในทั้งสองทิศทาง ($P < 0.05$) ขณะที่ภูเกิลแปลภาษาและไปตู้แปลภาษามีคุณภาพการแปลคำศัพท์ได้ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($P > 0.05$) สำหรับการแปลระดับวลีหรือประโยค ไม่พบปฏิสัมพันธ์ระหว่างเครื่องมือแปลและผู้ประเมินเช่นกัน แชตจีพีทีมีคะแนนคุณภาพการแปลสูงกว่าอีก 2 เครื่องมืออย่างมีนัยสำคัญทางสถิติในทุกด้านและทุกทิศทางการแปล ($P < 0.01$) ในทิศทางจีนเป็นไทย ภูเกิลแปลภาษามีคะแนนสูงกว่าไปตู้แปลภาษาอย่างมีนัยสำคัญทุกด้าน ($P < 0.05$) ส่วนในทิศทางไทยเป็นจีน ภูเกิลแปลภาษามีคะแนนสูงกว่าไปตู้แปลภาษาในด้าน adequacy, meaning และ severity ($P < 0.05$) แต่ไม่แตกต่างกันในด้าน fluency ($P = 0.424$) **สรุป:** แชตจีพีทีที่มีคุณภาพสูงที่สุดในการแปลคำ วลี หรือประโยคที่ใช้ในการชกประวัติเบื้องต้น ทั้งจากภาษาไทยเป็นภาษาจีนและจากภาษาจีนเป็นภาษาไทย อย่างไรก็ตามควรใช้เครื่องมือนี้เป็นเพียงเครื่องมือช่วยสื่อสารในเบื้องต้นเท่านั้น โดยบุคลากรทางการแพทย์ควรตรวจสอบข้อมูลสำคัญเพิ่มเติมเพื่อป้องกันความคลาดเคลื่อนที่อาจส่งผลกระทบต่อสุขภาพของผู้รับบริการ

คำสำคัญ: การแปลภาษาอัตโนมัติ การชกประวัติเบื้องต้น คุณภาพการแปล แชตจีพีที ไปตู้แปลภาษา ภูเกิลแปลภาษา

รับต้นฉบับ: 9 เม.ย. 2569, **ได้รับบทความฉบับปรับปรุง:** 31 พ.ค. 2569, **รับลงตีพิมพ์:** 4 มิ.ย. 2569

ผู้ประสานงานบทความ: วีรยุทธ์ เลิศนที สาขาสุขภาพดิจิทัล คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร เลขที่ 6 ถนนราชมรรคาใน ตำบลพระปฐมเจดีย์ อำเภอเมือง จังหวัดนครปฐม 73000 **E-mail:** lertnattee_v@su.ac.th

Evaluation of Translation Quality Between Chinese and Thai for Basic History Taking Using Google Translate, Baidu Translate, and ChatGPT

Papit Makkun¹, Verayuth Lertnattee²

¹Student in Master of Pharmacy Program in Health Informatics, Faculty of Pharmacy, Silpakorn University

²Department of Digital Health, Faculty of Pharmacy, Silpakorn University

Abstract

Objective: To assess and compare the translation quality of basic history-taking content from Chinese to Thai and from Thai to Chinese using Google Translate, Baidu Translate, and ChatGPT. **Methods:** This evaluative study began with testing three prompt formulations for ChatGPT in order to identify the most appropriate prompt for translating health-related text. Google Translate, Baidu Translate, and ChatGPT were then used to translate 75 disease- and symptom-related terms and 86 phrases or sentences relevant to basic history taking, neurology, gastroenterology, and reproductive systems. The translation materials were selected from a patient history-taking handbook. The translations were performed in both Chinese-to-Thai and Thai-to-Chinese directions. The translation quality was assessed by two experts at two levels: 1) the lexical level, which was evaluated using acceptability and compared across the three tools using Bayesian generalized linear mixed models; and 2) the phrasal/sentential level, which was evaluated using a 5-point Likert-scale questionnaire across four domains: adequacy, fluency (grammaticality and comprehensibility), meaning, and severity (impact on health). Comparisons of translation quality among the three tools were conducted using an ordinal logistic mixed-effects model. **Results:** At the lexical level, no statistically significant interaction was found between translation tools and raters. ChatGPT demonstrated the highest probability of acceptable translations in both the Chinese-to-Thai and Thai-to-Chinese directions, with acceptability rates of 99.98% and 99.96%, respectively. ChatGPT showed significantly better lexical translation quality than Google Translate and Baidu Translate in both directions ($P < 0.05$), whereas no statistically significant difference was observed between Google Translate and Baidu Translate ($P > 0.05$). At the phrasal/sentential level, no statistically significant interaction was found between translation tools and raters. ChatGPT achieved significantly higher translation quality scores than the other two tools across all domains and in both translation directions ($P < 0.01$). In the Chinese-to-Thai direction, Google Translate scored significantly higher than Baidu Translate in all domains ($P < 0.05$). In the Thai-to-Chinese direction, Google Translate scored significantly higher than Baidu Translate in adequacy, meaning, and severity ($P < 0.05$), but no significant difference was found in fluency ($P = 0.424$). **Conclusion:** ChatGPT demonstrated the highest translation quality for words, phrases, or sentences used in basic history taking, in both Thai-to-Chinese and Chinese-to-Thai directions. However, it should be used only as a communication aid. Healthcare professionals should further verify critical information to prevent inaccuracies that may affect patient health.

Keywords: machine translation, basic history taking, translation quality, ChatGPT, Baidu Translate, Google Translate

บทนำ

ประเทศไทยและสาธารณรัฐประชาชนจีนมีความสัมพันธ์ทางการทูตที่แน่นแฟ้นและยาวนานมากกว่า 40 ปี นับตั้งแต่การสถาปนาความสัมพันธ์อย่างเป็นทางการในปี พ.ศ. 2518 ความร่วมมือระหว่างสองประเทศครอบคลุมหลากหลายมิติ ทั้งด้านเศรษฐกิจ การค้า การลงทุน การศึกษา และการท่องเที่ยว ส่งผลให้จีนกลายเป็นหนึ่งในประเทศคู่ค้าทางเศรษฐกิจที่สำคัญที่สุด รวมถึงเป็นตลาดนักท่องเที่ยวหลักของประเทศไทยมาอย่างต่อเนื่อง (1) ในช่วง 10 เดือนแรกของปี พ.ศ. 2568 จำนวนนักท่องเที่ยวจีนมีมากเป็นอันดับที่ 4 (ร้อยละ 13 ของจำนวนนักท่องเที่ยวต่างชาติในไทย) รองจากชาวญี่ปุ่น สหรัฐอเมริกา และสิงคโปร์ การลงทุนมีมูลค่าประมาณ 25,404 ล้านบาท (2) ในเดือนแรกของปี พ.ศ. 2569 พบว่า นักลงทุนชาวจีนเป็นธุรกิจต่างชาติในไทยที่มีจำนวนมากที่สุด (3) แสดงให้เห็นว่ามีการลงทุนของกลุ่มนักลงทุนจากสาธารณรัฐประชาชนจีนอย่างต่อเนื่อง สถิติการทำงานของคนต่างด้าวประจำเดือนพฤศจิกายน พ.ศ. 2568 รายงานว่า จำนวนแรงงานชาวจีนที่เข้ามาทำงานในประเทศไทยนั้นมีปริมาณมากเช่นกัน (4) จำนวนนักศึกษาชาวจีนที่เข้ามาศึกษาต่อในประเทศไทยยังมีแนวโน้มเพิ่มสูงขึ้นอย่างต่อเนื่องตั้งแต่ปี พ.ศ. 2562 เป็นต้นมา (5) แสดงให้เห็นถึงการเข้ามาอยู่อาศัยในประเทศไทยของชาวจีนในหลากหลายสถานะ ไม่ว่าจะเป็นนักลงทุน แรงงาน หรือนักศึกษา ซึ่งล้วนมีความจำเป็นต้องเข้าถึงบริการด้านสุขภาพในชีวิตประจำวัน

ภายหลังการคลี่คลายของสถานการณ์การแพร่ระบาดของโรคโควิด-19 ประเทศไทยได้ต้อนรับนักท่องเที่ยวจีนจำนวนมาก จากสถิตินักท่องเที่ยวต่างชาติที่เดินทางเข้าประเทศไทย ในปี พ.ศ. 2567 มีนักท่องเที่ยวชาวจีนเดินทางมาประเทศไทยประมาณ 6.7 ล้านคน (6) และในปี 2568 ประมาณ 4.4 ล้านคน (7) ทั้งนี้แม้ในปี พ.ศ. 2568 จะมีจำนวนนักท่องเที่ยวลดลง แต่จากการเปรียบเทียบสถิติของปี พ.ศ. 2568 และ พ.ศ. 2569 พบว่า ใน 4 เดือนแรกของปี พ.ศ. 2569 มีจำนวนนักท่องเที่ยวจีนที่เดินทางมายังประเทศไทยมากกว่าปี พ.ศ. 2568 คิดเป็นร้อยละ 15 (8) แสดงถึงแนวโน้มนักท่องเที่ยวจีนที่จะเดินทางมายังประเทศไทยมากขึ้น นอกจากการท่องเที่ยวทั่วไปแล้ว การท่องเที่ยวเชิงการแพทย์ยังได้รับความนิยมในกลุ่มนักท่องเที่ยวชาวจีน เนื่องจากประเทศไทยมีมาตรฐานการให้บริการทางการแพทย์ที่มีคุณภาพและ

ค่าใช้จ่ายเหมาะสม และผู้รับบริการดังกล่าวยังสามารถเดินทางไปยังแหล่งท่องเที่ยวอื่นที่น่าสนใจได้ด้วย ธุรกิจสุขภาพของไทยได้รับการสนับสนุนจากภาครัฐด้วยนโยบายผลักดันให้ไทยเป็นศูนย์กลางสุขภาพนานาชาติ อีกทั้งผู้ประกอบการโรงพยาบาลเอกชนต่างเร่งขยายขีดความสามารถในการให้บริการเพื่อดึงดูดนักท่องเที่ยวเชิงการแพทย์ทั่วโลก (9)

เมื่อมีชาวจีนและนักท่องเที่ยวจีนเข้ามาในประเทศไทยมากขึ้น โอกาสที่จะเกิดภาวะเจ็บป่วยและต้องพึ่งพา ระบบสาธารณสุขไทยก็เพิ่มขึ้น อย่างไรก็ตาม ปัญหาสำคัญในการเข้าถึงบริการด้านสุขภาพของชาวจีนในประเทศไทยคือ อุปสรรคด้านการสื่อสาร เนื่องจากผู้มารับบริการชาวจีนส่วนใหญ่สามารถสื่อสารได้เฉพาะภาษาจีน ในขณะที่บุคลากรทางการแพทย์ส่วนใหญ่ไม่สามารถสื่อสารภาษาจีนได้ โดยขั้นตอนการซักประวัติผู้ป่วยถือเป็นขั้นตอนแรกและเป็นขั้นตอนที่มีความสำคัญอย่างยิ่ง เนื่องจากจะนำไปสู่การตรวจร่างกาย การวินิจฉัยโรค และการวางแผนการรักษาอย่างถูกต้อง มีประสิทธิภาพ และปลอดภัย (10)

ปัจจุบันมีการนำเครื่องมือแปลภาษามาช่วยในการสื่อสารระหว่างบุคลากรทางการแพทย์และผู้มารับบริการชาวจีน ซึ่งเครื่องมือแปลภาษาที่เป็นที่นิยมในประเทศไทยเนื่องจากเข้าถึงง่ายและไม่มีความจำเป็นต้องใช้การบริการคือ กูเกิลแปลภาษา (Google Translate) (11) การศึกษาในอดีตได้เปรียบเทียบการแปลภาษาไทยเป็นจีนโดยใช้กูเกิลแปลภาษา ไปตู้แปลภาษา (Baidu Translate) และแชตจีพีที (ChatGPT) (12) ในส่วนของประเทศจีนเครื่องมือแปลภาษาที่เป็นที่รู้จักดี คือ ไปตู้แปลภาษา (13) และ Youdao Translate (14) แต่เครื่องมือที่เป็นที่นิยมและมีการใช้มายาวนานกว่าคือ ไปตู้แปลภาษา นอกจากนี้ยังมีปัญญาประดิษฐ์เชิงสร้างสรรค์ (generative AI) ที่ได้รับความนิยมในขณะนี้คือ แชตจีพีที (15) ที่มีความสามารถในการรับรู้บริบทของการสนทนา ซึ่งมีความสามารถในการแปลภาษาด้วย

กูเกิลแปลภาษาและไปตู้แปลภาษานั้นเป็นเครื่องมือแปลภาษาแบบ neural machine translation โดยมีหลักการทำงานที่ใช้เครือข่ายตัวเข้ารหัสแปลงประโยคต้นทางเป็นเวกเตอร์ที่มีค่าจริง จากนั้นส่งต่อลำดับเวกเตอร์นั้นไปประมวลผลแสดงความหมายก่อนที่จะเครือข่ายตัวถอดรหัสจะสร้างประโยคในภาษาเป้าหมาย (16,17) ซึ่งจะแตกต่างจากแชตจีพีทีที่เป็นปัญญาประดิษฐ์เชิงสร้างสรรค์ที่อาศัย

โมเดลภาษาขนาดใหญ่ ซึ่งได้รับการฝึกฝนจากข้อมูลจำนวนมาก ทำให้สามารถเข้าใจและสร้างภาษามนุษย์ได้อย่างเป็นธรรมชาติ รวมถึงใช้เทคโนโลยีการเรียนรู้เชิงลึกในการเรียนรู้และสร้างผลงานที่มีความสร้างสรรค์ขึ้นมาได้ (18,19) การนำแชตจีพีทีมาใช้ในการแปลภาษาทำได้โดยใช้คำสั่งแปลข้อความที่ต้องการไปเป็นภาษาปลายทาง (20)

อย่างไรก็ตาม เครื่องมือแปลภาษาและปัญญาประดิษฐ์เชิงสร้างสรรค์ของกลุ่มนี้ ควรมีการประเมินคุณภาพในการแปลภาษาก่อน เนื่องจากข้อมูลที่ได้จากการแปลภาษา โดยเฉพาะที่ใช้ในการซักประวัติมีความสำคัญอย่างมากในการตรวจร่างกาย การวินิจฉัยโรค และการวางแผนการรักษา ซึ่งหากเกิดความคลาดเคลื่อนอาจส่งผลกระทบต่อความปลอดภัยของผู้รับบริการชาวจีนได้ จากการทบทวนวรรณกรรมไม่พบการศึกษาคุณภาพในการแปลภาษาของกูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีที ในคู่ภาษาจีนกับภาษาไทยเพื่อการสื่อสารในการซักประวัติเบื้องต้น พบเพียงการศึกษาประสิทธิภาพในเชิงภาษาศาสตร์ในการแปลภาษาไทยเป็นภาษาจีนที่ไม่เกี่ยวข้องกับทางการแพทย์และสุขภาพของกัลยา ขาวบ้านแพ้ว (21) ที่ศึกษาประสิทธิภาพการแปลนิตยสารไทยแอร์เอเชีย โดยพบว่า กูเกิลแปลภาษามีประโยชน์ที่แปลภาษาได้คุณภาพระดับดีร้อยละ 30 ปานกลางร้อยละ 60 และควรปรับปรุงร้อยละ 10 ในส่วนของไปตู้แปลภาษามีประโยชน์ที่แปลภาษาได้คุณภาพระดับดีร้อยละ 3.89 ปานกลางร้อยละ 46.67 และควรปรับปรุงร้อยละ 49.44 นอกจากนี้ยังมีการศึกษาของ Zuo และคณะ (12) ที่ศึกษาการประยุกต์ใช้แชตจีพีทีในการแปลวรรณกรรมจากภาษาไทยเป็นภาษาจีน โดยวัดด้วยคะแนนความคล้ายคลึง (similarity score) โดย 0 แสดงถึง ความไม่คล้ายคลึงของประโยคภาษาต้นทางและปลายทาง และ 1 แสดงถึงความเหมือนกันโดยสมบูรณ์ ซึ่งผลการศึกษาพบว่าการแปลของกูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีทีมีคะแนนความคล้ายคลึงเท่ากับ 0.25, 0.23 และ 0.33 ตามลำดับ

นอกจากนี้ยังพบการศึกษาคุณภาพของกูเกิลแปลภาษาและแชตจีพีทีในการแปลวลีหรือประโยคระหว่างภาษาจีนกับภาษาอื่นในบริบทด้านสุขภาพ ดังนี้ ในส่วนของกูเกิลแปลภาษา Chen และคณะ (22) ได้ประเมินความถูกต้องของการใช้กูเกิลแปลภาษาในการแปลแผนปฏิบัติการดูแลตนเองของผู้ป่วยเบาหวาน โดยแปลจากภาษาอังกฤษ

ไปเป็นภาษาสเปนและภาษาจีน และประเมินความถูกต้องด้วยแบบสอบถามที่เป็นมาตรวัดลิเคิร์ต 5 ระดับใน 4 หัวข้อ คือ fluency (สามารถอ่านได้ ไวยากรณ์ถูกต้อง และเข้าใจได้), adequacy (ข้อมูลตรงตามต้นฉบับ), meaning (ความหมายตรงกับต้นฉบับ) และ severity (การแปลส่งผลกระทบต่อผลลัพธ์ด้านสุขภาพของผู้ป่วย) พบว่าคะแนนเฉลี่ยอยู่ที่ 1.67 ถึง 5 ส่วนการศึกษาของ Taira และคณะ (23) ได้ประเมินความถูกต้องของกูเกิลแปลภาษาจากการแปลเอกสารคำแนะนำหลังจากออกจากห้องฉุกเฉิน ภาษาอังกฤษไปเป็น 7 ภาษารวมถึงภาษาจีน ในส่วนของภาษาจีนมีคะแนน fluency, adequacy, meaning และ severity อยู่ที่ 4.1, 4.5, 4.1 และ 4.1 ตามลำดับ นอกจากนี้ยังพบอัตราความถูกต้อง (accuracy rate) ซึ่งคิดจากข้อความที่ความหมายยังคงอยู่หารด้วยข้อความทั้งหมดที่แปล มีค่าเท่ากับร้อยละ 81.7 ขณะที่ Khanna และคณะ (24) ประเมินแชตจีพีทีในการแปลรายงานรังสีวิทยาจากภาษาอังกฤษไปยังอีก 5 ภาษา โดยใช้คำสั่ง 2 แบบคือคำสั่งให้แปลโดยตรงและคำสั่งให้เรียบเรียงรายงานที่แปลใหม่ด้วยภาษาที่คนทั่วไปเข้าใจ แล้วให้ผู้ประเมินประเมินด้วยมาตรวัดลิเคิร์ต 5 ระดับในประเด็นความแม่นยำในการแปล ความชัดเจนของรายงานที่ย่อลง และความเข้าใจของผู้ป่วย ซึ่งในส่วนของภาษาจีนพบว่า โดยรวมผู้ประเมินมีความเห็นด้วยและเห็นด้วยอย่างยิ่งในระดับสูง ต่อคุณภาพการแปลของแชตจีพีที

จากการทบทวนวรรณกรรมที่ผ่านมา ยังไม่พบงานวิจัยที่มุ่งประเมินคุณภาพของเครื่องมือที่ใช้ในการแปลภาษาระหว่างภาษาจีนและภาษาไทยที่เกี่ยวข้องกับการซักประวัติผู้ป่วยโดยตรง การศึกษานี้จึงมีวัตถุประสงค์เพื่อประเมินคุณภาพในการแปลภาษาระหว่างภาษาจีนและภาษาไทยสำหรับการสื่อสารเพื่อซักประวัติเบื้องต้นของกูเกิลแปลภาษา ไปตู้แปลภาษา และแชตจีพีที รวมถึงเพื่อเปรียบเทียบคุณภาพของเครื่องมือทั้งสามว่ามีความแตกต่างกันเพียงใด

วิธีการวิจัย

การศึกษานี้เป็นการวิจัยเชิงประเมินประสิทธิภาพที่ผ่านการพิจารณาจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยศิลปากร (เลขที่โครงการ REC 68.0825-146-7437) โดยมีรายละเอียดการดำเนินการวิจัยดังนี้

การคัดเลือกข้อความต้นฉบับสำหรับแปล

ผู้วิจัยได้คัดเลือกคำ วลี หรือประโยคที่เกี่ยวข้องกับการสื่อสารเพื่อชกประวัติเบื้องต้นทั้งภาษาจีนและภาษาไทยจากคู่มือการชกประวัติผู้รับบริการ 3 ภาษา (ไทย อังกฤษ และจีน) (26) โดยสิ่งที่คัดเลือกจำแนกเป็น 2 ส่วนหลัก ได้แก่ ส่วนของคำศัพท์เกี่ยวกับโรคและอาการจำนวน 75 คำ และส่วนของวลีหรือประโยคจำนวน 86 วลีหรือประโยค ซึ่งแบ่งเป็น 4 ประเภทย่อย ได้แก่ วลีหรือประโยคที่เกี่ยวข้องกับการชกประวัติเบื้องต้น วลีหรือประโยคที่เกี่ยวข้องกับระบบประสาท วลีหรือประโยคที่เกี่ยวข้องกับระบบทางเดินอาหาร และวลีหรือประโยคที่เกี่ยวข้องกับระบบสืบพันธุ์

การคัดเลือกคำสั่งการแปลสำหรับแชตจีพีที

การศึกษานำวลีหรือประโยคที่คัดเลือกจากคู่มือการชกประวัติผู้รับบริการ 3 ภาษา (ไทย อังกฤษ และจีน) จำนวน 12 วลีหรือประโยคโดยใช้การสุ่มแบบง่าย แล้วนำมาแปลผ่านแชตจีพีที (GPT-5.0) ผ่านช่องทางเว็บอินเทอร์เฟซ และควบคุมให้แปลแบบ session เดียวต่อเนื่อง ณ วันที่ 17 ตุลาคม 2568 ด้วยคำสั่งที่แตกต่างกัน 3 คำสั่ง เพื่อทดสอบคุณภาพของคำสั่งทั้งสามว่ามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ และเลือกคำสั่งที่เหมาะสมที่สุดสำหรับการนำไปใช้ในการแปล

ในขั้นตอนประเมินคุณภาพการแปลจริง การกำหนดคำสั่งแบ่งออกเป็นคำสั่งพื้นฐาน 1 คำสั่ง ซึ่งเป็นการสั่งการแปลโดยไม่มีการกำหนดบริบทเพิ่มเติม และคนทั่วไปมีโอกาสใช้คำสั่งนี้มากที่สุด ส่วนอีก 2 คำสั่งเป็นคำสั่งกำหนดบทบาทซึ่งมีการระบุบทบาทของผู้เชี่ยวชาญด้านสุขภาพและความเป็นเจ้าของภาษา โดยคำสั่งที่ 2 และ 3 มีความหมายเหมือนกัน แต่มีความแตกต่างกันเพียงโครงสร้างทางไวยากรณ์ คำสั่งทั้งสามมีดังนี้ 1) translate "....." into Chinese/Thai. 2) translate "....." into Chinese/Thai, act as health professional who is a native speaker of the target language. The goal is to translate text fluently and adequately. 3) As a health professional who is a native speaker of the target language, translate "....." into Chinese/Thai. Provide fluent and adequate translated text. การแปลมีทั้งแปลจากภาษาจีนเป็นภาษาไทย และจากภาษาไทยเป็นภาษาจีน ตามลำดับ

ผลการแปลถูกประเมินโดยผู้ประเมินจำนวน 2 ท่าน ซึ่งเป็นคนไทยที่สำเร็จการศึกษาในระดับปริญญาตรี

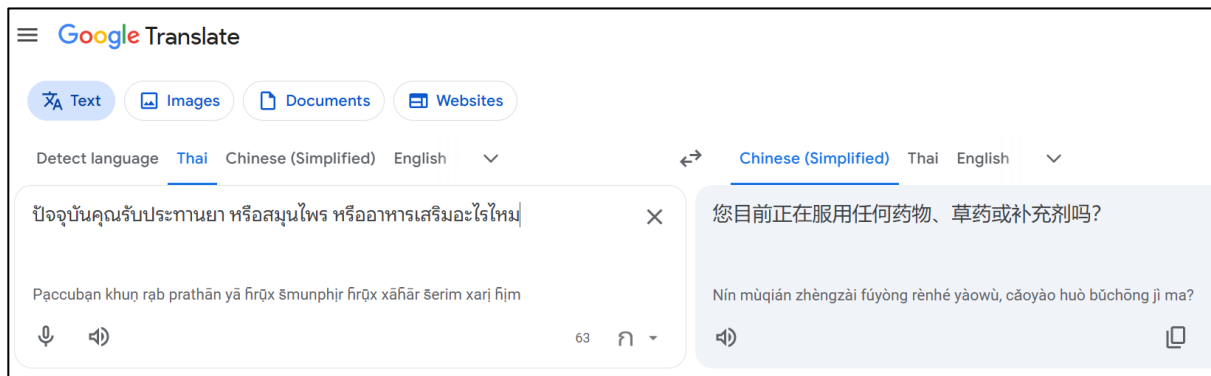
ขึ้นไปสาขาภาษาจีน การคัดเลือกผู้ประเมินใช้วิธีการเลือกแบบเฉพาะเจาะจงโดยพิจารณาจากคุณสมบัติที่กำหนดไว้ และติดต่อเชิญผู้ที่มีคุณสมบัติดังกล่าวเข้าร่วมโครงการวิจัยโดยตรง การประเมินผลใช้แบบสอบถามที่เป็นมาตรวัดลิเคิร์ต 5 ระดับ แบบประเมินประกอบด้วย 2 หัวข้อ ได้แก่ adequacy (การแปลข้อมูลได้ครบถ้วนตรงตามต้นฉบับ) และ fluency (การแปลให้เป็นไปตามหลักไวยากรณ์และสามารถเข้าใจได้ง่าย) ทั้งนี้ การประเมินดำเนินการในลักษณะปกปิด โดยผู้ประเมินไม่ทราบว่าภาษาปลายทางที่ได้รับการแปลถูกแปลมาจากคำสั่งใด

การศึกษาใช้สถิติ ordinal logistic mixed-effects model เพื่อทำนายระดับ adequacy และ fluency โดยมีตัวแปรอิสระหลักในการวิเคราะห์ ได้แก่ คำสั่ง (prompt) ที่ใช้ และมีตัวแปรสุ่ม ได้แก่ ผู้ประเมินและวลีหรือประโยคที่ได้รับการสุ่มมา ทั้งนี้ การวิเคราะห์ข้อมูลจากการแปลจากภาษาไทยเป็นภาษาจีน และการวิเคราะห์ข้อมูลจากการแปลจากภาษาจีนเป็นภาษาไทย แยกวิเคราะห์อย่างเป็นอิสระต่อกัน การวิเคราะห์ทำเพื่อทดสอบว่าคำสั่งทั้งสามมีคุณภาพแตกต่างกันอย่างมีนัยสำคัญหรือไม่ เพื่อเลือกคำสั่งสำหรับนำไปใช้ในการทดสอบคุณภาพการแปลของแชตจีพีทีในขั้นตอนต่อไป

การประเมินคุณภาพการแปลของเครื่องมือ

การประเมินคุณภาพการแปลดำเนินการโดยผู้ประเมิน 2 ท่านที่มีคุณสมบัติ ดังนี้ 1) เป็นชาวไทยที่จบการศึกษาในสาขาที่เกี่ยวข้องกับสุขภาพในระดับปริญญาตรีเป็นอย่างน้อย และ 2) เคยผ่านการสอบวัดระดับความรู้ภาษาจีน (HSK) ตั้งแต่ระดับ 5 ขึ้นไป ผู้ประเมิน 2 ท่านในงานวิจัยนี้เป็นแพทย์ที่มีประสบการณ์การตรวจรักษาผู้ป่วยทางคลินิก

ผู้ประเมินได้รับคำและวลีหรือประโยคภาษาต้นทางที่เป็นต้นฉบับ และคำและวลีหรือประโยคภาษาปลายทางที่ได้รับการแปลผ่าน 3 เครื่องมือ ได้แก่ ภูเก็ต แปลภาษา ไปดูแปลภาษา และแชตจีพีที (GPT-5.0) การแปลของแปลภาษาและไปดูแปลภาษาเป็นการแปลแบบ session เดียวต่อเนื่องผ่านช่องทางเว็บอินเทอร์เฟซ (รูปที่ 1) ส่วนการแปลของแชตจีพีทีเป็นการแปลแบบ session เดียวต่อเนื่องผ่านช่องทาง API โดยกำหนดค่า temperature = 1.0 ณ วันที่ 31 ตุลาคม 2568 ทั้งนี้ วิธีการที่กล่าวมาใช้กับทั้งการแปลจากภาษาจีนเป็นภาษาไทย และจากภาษาไทยเป็นภาษาจีน



ก. หน้าจอเว็บไซต์กูเกิลแปลภาษา (<https://translate.google.com/>)



ข. หน้าจอเว็บไซต์ไปตู้แปลภาษา (<https://fanyi.baidu.com/>)

รูปที่ 1. หน้าจอเว็บไซต์กูเกิลแปลภาษาและไปตู้แปลภาษา

การประเมินจะดำเนินการในลักษณะปกปิดข้อมูล โดยผู้ประเมินจะไม่ทราบว่ภาษาปลายทางที่ได้รับการแปล ถูกแปลมาจากแอปพลิเคชันใดเพื่อป้องกันอคติในการ ประเมิน การประเมินแบ่งออกเป็น 2 ส่วน ได้แก่ 1) การ ประเมินคำศัพท์ เป็นการประเมินค่าการยอมรับได้ (27) หมายถึง ความสามารถในการแปลให้เข้าใจได้และสามารถ นำไปใช้ในบริบททางการแพทย์ได้ โดยให้ผู้ประเมินตัดสิน ว่าคำแปลนั้นยอมรับได้หรือยอมรับไม่ได้ และ 2) การ ประเมินวลีหรือประโยค โดยใช้แบบสอบถามแบบมาตรวัดลิ เคิร์ท 5 ระดับ ซึ่งประเมินใน 4 ด้าน ได้แก่ adequacy, fluency, meaning และ severity (22,25) adequacy คือ ความสามารถในการแปลเนื้อหาและสารสนเทศครบถ้วน ตรงตามต้นฉบับ; fluency คือ ความสามารถในการแปลให้ เป็นไปตามหลักไวยากรณ์และสามารถเข้าใจได้อย่าง ราบรื่น; meaning: คือ ความสามารถในการแปลให้ได้ ความหมายที่ถูกต้องและตรงตามต้นฉบับ; และ severity คือ ความคลาดเคลื่อนในการแปลที่อาจส่งผลกระทบต่อสุขภาพของผู้ป่วย ความหมายของระดับคะแนน 1-5 ใน แต่ละด้านที่ประเมินแสดงอยู่ในตารางที่ 1

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลใช้ภาษา R โดยใช้เครื่องมือ R Studio เวอร์ชัน 2026.01.0 Build 392 บนระบบปฏิบัติการ Windows 10 Pro การวิเคราะห์แบ่งออกเป็น 2 ส่วน ได้แก่

1) การวิเคราะห์คุณภาพการแปลระดับคำศัพท์ วิเคราะห์โดยใช้ค่ายอมรับได้ และทดสอบความสอดคล้อง ของผู้ประเมินด้วย weighted Cohen's kappa coefficient จากนั้น ทดสอบสมมติฐานด้วยสถิติ Bayesian generalized linear mixed models โดยมีตัวแปรอิสระหลัก ได้แก่ เครื่องมือแปลภาษา และมีตัวแปรสุ่ม ได้แก่ ผู้ประเมินและ คำศัพท์ที่ได้รับการคัดเลือกมา ทั้งนี้ การวิเคราะห์ระหว่าง ข้อมูลการแปลจากภาษาไทยเป็นภาษาจีน และข้อมูลการ แปลจากภาษาจีนเป็นภาษาไทย แยกวิเคราะห์อย่างเป็น อิสระต่อกัน การวิเคราะห์กำหนดลักษณะการกระจายตัว แบบ weakly informative เพื่อเพิ่มประสิทธิภาพในการ ประมวลผลและลดปัญหาการไม่ลู่เข้าของโมเดล โดยผู้วิจัย กำหนดให้พารามิเตอร์อิทธิพลคงที่ (fixed effects) มีการ กระจายตัวแบบปกติ normal(0, 2.5) และกำหนดให้ส่วน เบี่ยงเบนมาตรฐานของอิทธิพลสุ่ม (random effects) มีการ

ตารางที่ 1. เกณฑ์การประเมินคุณภาพของการแปลและความหมายที่ใช้ในงานวิจัย

ระดับ	adequacy	fluency	meaning	severity
1	ข้อมูลไม่ตรงตามต้นฉบับ	ไม่เป็นไปตามหลักไวยากรณ์และไม่สามารถเข้าใจได้	ความหมายที่ได้ไม่เป็นไปตามต้นฉบับ	ความคลาดเคลื่อนทำให้เกิดอันตรายต่อสุขภาพ
2	ข้อมูลครบถ้วนตรงตามต้นฉบับน้อย	ผิดพลาดทางหลักไวยากรณ์สูง	ความหมายที่ได้เป็นไปตามต้นฉบับน้อย	ความคลาดเคลื่อนทำให้การดูแลผู้ป่วยบกพร่อง
3	ข้อมูลครบถ้วนตรงตามต้นฉบับปานกลาง	ผิดพลาดทางหลักไวยากรณ์สูง แต่เข้าใจได้	ความหมายที่ได้เป็นไปตามต้นฉบับปานกลาง	ความคลาดเคลื่อนทำให้การดูแลผู้ป่วยล่าช้า
4	ข้อมูลครบถ้วนตรงตามต้นฉบับมาก	ผิดพลาดทางหลักไวยากรณ์น้อย	ความหมายที่ได้เป็นไปตามต้นฉบับมาก	ความคลาดเคลื่อนส่งผลกระทบต่อสุขภาพไม่ชัดเจน
5	ข้อมูลครบถ้วนตรงตามต้นฉบับอย่างสมบูรณ์	เป็นไปตามหลักไวยากรณ์	ความหมายที่ได้เป็นไปตามต้นฉบับอย่างสมบูรณ์	ความคลาดเคลื่อนไม่ส่งผลกระทบต่อสุขภาพ

กระจายตัวแบบคอชี cauchy(0, 2.5) และอาศัยการประมาณค่าจาก posterior distribution เพื่อคำนวณหาความน่าจะเป็นและ odds ratio ในการทดสอบนัยสำคัญทางสถิติ ในการนี้ ผู้วิจัยเลือกใช้ค่า P ที่ระดับ 0.05 เป็นเกณฑ์ การเปรียบเทียบความแตกต่างของเครื่องมือทั้งสาม โดยไม่ได้ปรับค่า P เนื่องจากงานวิจัยนี้เป็นการศึกษาเชิงบุกเบิกเพื่อประเมินและเปรียบเทียบคุณภาพเบื้องต้น การใช้สถิติปรับแก้ค่า P ที่เข้มงวดเกินไปอาจเพิ่มความเสี่ยงในการเกิดความคลาดเคลื่อนประเภทที่ 2 ซึ่งบดบังความแตกต่างของเครื่องมือที่มีนัยสำคัญทางคลินิกหรือทางปฏิบัติได้

2) การวิเคราะห์คุณภาพการแปลระดับวลีหรือประโยค เป็นการวิเคราะห์ใน 4 ด้าน ได้แก่ adequacy, fluency, meaning และ severity การวิเคราะห์เริ่มด้วยการทดสอบความสอดคล้องของผู้ประเมินด้วย weighted Cohen's kappa coefficient จากนั้นหาค่าสถิติเชิงพรรณนา และทดสอบสมมติฐานด้วยสถิติ ordinal logistic mixed-effects model เพื่อเปรียบเทียบความแตกต่างของเครื่องมือทั้งสาม โดยมีตัวแปรอิสระหลัก ได้แก่ เครื่องมือแปลภาษา และมีตัวแปรสุ่ม ได้แก่ ผู้ประเมินและวลีหรือประโยคที่ได้รับการคัดเลือกมา ทั้งนี้ การวิเคราะห์ระหว่างข้อมูลการแปลจากภาษาไทยเป็นภาษาจีน และข้อมูลการแปลจากภาษาจีนเป็นภาษาไทย แยกวิเคราะห์อย่างเป็นอิสระต่อกัน

การศึกษานี้แปลผลขนาดของค่า kappa ตามเกณฑ์ของ Landis และ Koch (28) คือ ค่า 0.01 – 0.20, 0.21 – 0.40, 0.41 – 0.60, 0.61 – 0.80 และ 0.81 – 1.00 บ่งบอกถึงระดับความสอดคล้องระหว่างการประเมินในระดับน้อย พอใช้ ปานกลาง ดี และดีเยี่ยม

ผลการวิจัย

คำสั่งที่เหมาะสมที่สุดสำหรับแชตจีพีที

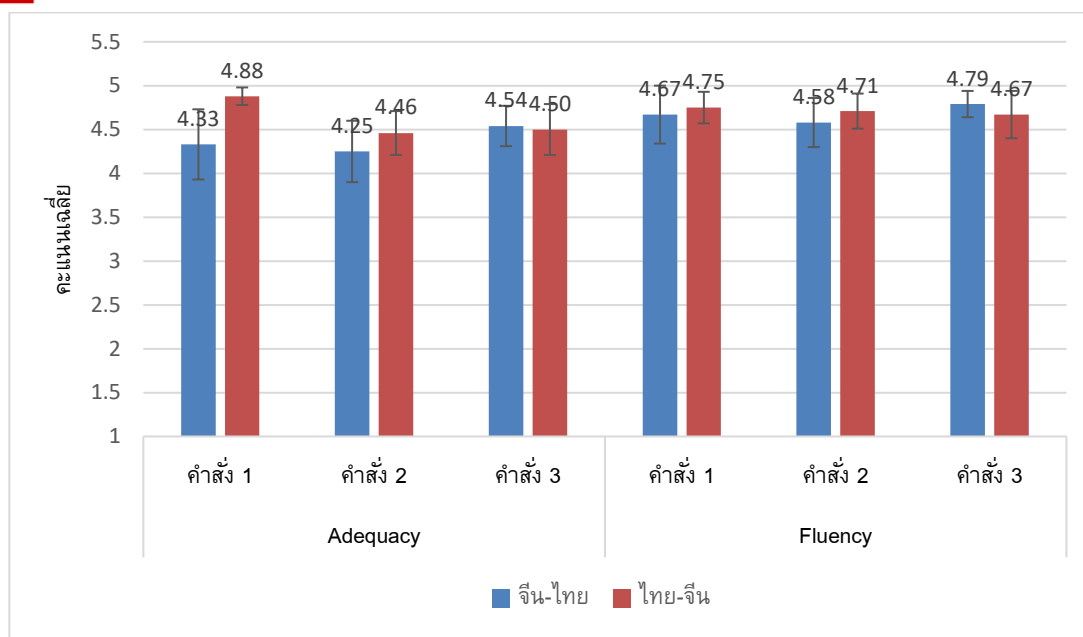
การแปลจากภาษาจีนเป็นภาษาไทย

ผลการเปรียบเทียบคำสั่ง (prompt) ของแชตจีพีทีในการแปลวลีหรือประโยคภาษาจีนเป็นภาษาไทยแสดงอยู่ในรูปที่ 2 พบว่า คำสั่งที่ 3 “As a health professional who is a native speaker of the target language, translate “.....” into Thai. Provide fluent and adequate translated text.” มีค่าเฉลี่ยของ adequacy และ fluency สูงสุดและมีความคลาดเคลื่อนมาตรฐาน (standard error: SE) ต่ำสุดเมื่อเทียบกับอีก 2 คำสั่ง โดยมีค่าเท่ากับ 4.54 (SE = 0.23) และ 4.79 (SE = 0.15) ตามลำดับ

การทดสอบด้วย ordinal logistic mixed-effects model ไม่พบปฏิสัมพันธ์ระหว่างคำสั่งที่ใช้และผู้ประเมินอย่างมีนัยสำคัญทางสถิติทั้งในการประเมิน adequacy (P = 0.549) และ fluency (P = 0.281) การวิเคราะห์อิทธิพลหลักพบว่า คำสั่งไม่ส่งผลต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญทั้งด้าน adequacy (P = 0.712) และ fluency (P = 0.517) ในส่วนของผู้ประเมินไม่ส่งผลต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญในด้าน fluency (P = 0.187) แต่ส่งผลต่อคุณภาพอย่างมีนัยสำคัญในด้าน adequacy (P = 0.005)

การแปลจากภาษาไทยเป็นภาษาจีน

ผลการเปรียบเทียบคำสั่งของแชตจีพีทีในการแปลวลีหรือประโยคจากภาษาไทยเป็นภาษาจีน พบว่า คำสั่งที่ 1 “translate “.....” into Chinese” มีค่าเฉลี่ยสูงสุดและ SE ต่ำสุดเมื่อเทียบกับอีก 2 คำสั่งทั้งในด้าน adequacy



รูปที่ 2. คะแนนเฉลี่ยของคุณภาพการแปลด้าน adequacy และ fluency (เต็ม 5 คะแนน) เมื่อจำแนกตามคำสั่งที่ใช้ในแชตจีพีทีและทิศทางการแปล (จากภาษาจีนเป็นภาษาไทย หรือจากภาษาไทยเป็นภาษาจีน)

หมายเหตุ: คำสั่งที่ 1: translate “.....” into Chinese/Thai; คำสั่งที่ 2: translate “.....” into Chinese/Thai, act as health professional who is a native speaker of the target language. The goal is to translate text fluently and adequately; และ คำสั่งที่ 3: As a health professional who is a native speaker of the target language, translate “.....” into Chinese/Thai. Provide fluent and adequate translated text.

และ fluency โดยมีค่าเฉลี่ยเท่ากับ 4.88 (SE = 0.10) และ 4.75 (SE = 0.18) ตามลำดับ (รูปที่ 2) นอกจากนี้ผลการทดสอบด้วย ordinal logistic mixed-effects model พบว่า คำสั่งและผู้ประเมินไม่มีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติในทั้งการประเมิน adequacy ($P = 0.995$) และ fluency ($P = 0.363$) ผลการวิเคราะห์อิทธิพลหลักพบว่า คำสั่งไม่ส่งผลต่อคะแนนคุณภาพการแปลทั้งในด้าน adequacy ($P = 0.199$) และ fluency ($P = 0.990$) ในส่วนของผู้ประเมินพบว่าไม่ส่งผลต่อคุณภาพการแปลอย่างมีนัยสำคัญในด้าน adequacy ($P = 0.267$) แต่ส่งผลต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญในด้าน fluency ($P = 0.003$)

คุณภาพของการแปลระดับคำศัพท์

ผลการประเมินคุณภาพของการแปลระดับคำศัพท์ระหว่างผู้ประเมิน 2 ท่านพบว่ามีความสอดคล้องกันในระดับดี (Kappa = 0.75)

การแปลคำศัพท์จากจีนเป็นไทย

สำหรับคุณภาพการแปลระดับคำศัพท์จากภาษาจีนเป็นภาษาไทย พบว่า การแปลจากแชตจีพีทีที่มี

โอกาสที่จะได้คำแปลที่ยอมรับได้สูงสุด (ร้อยละ 99.98; 95%CI: 99.43-100.00) ตามด้วยคำแปลจากกูเกิลแปลภาษา (ร้อยละ 98.46; 95%CI: 95.64-99.92) และไปตู้แปลภาษา (ร้อยละ 96.10; 95%CI: 90.81-99.57)

ผลการวิเคราะห์ด้วย Bayesian generalized linear mixed models พบว่า ไม่มีปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมินอย่างมีนัยสำคัญทางสถิติ ($P > 0.05$) ผู้ประเมินไม่ส่งผลต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญ ($P = 0.152$) (ตารางที่ 2 ก และ ข)

การเปรียบเทียบรายคู่พบว่า แชตจีพีทีที่มีคะแนนการยอมรับได้สูงกว่าไปตู้แปลภาษาและกูเกิลแปลภาษาอย่างมีนัยสำคัญทางสถิติ ($P < .001$ และ $p = 0.010$ ตามลำดับ) ขณะที่ไปตู้แปลภาษาและกูเกิลแปลภาษามีคะแนนยอมรับได้ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($P = 0.122$) (ตารางที่ 2 ค)

การแปลคำศัพท์จากไทยเป็นจีน

ผลการประเมินคุณภาพการแปลระดับคำศัพท์จากภาษาไทยเป็นภาษาจีน พบว่า เครื่องมือแปลทั้ง 3 ให้คำแปลที่ยอมรับได้ในระดับที่ใกล้เคียงกันมาก โดยมีค่าการ

ตารางที่ 2. ผลของการวิเคราะห์ด้วย Bayesian generalized linear mixed models เพื่อทำนายการยอมรับได้ของผลการแปลระดับคำศัพท์จากภาษาจีนเป็นภาษาไทย

ก. ผลการทดสอบอิทธิพลหลักของผู้ประเมินผลของการแปลระดับคำศัพท์

รายการประเมิน	ค่า odds ratio (OR)	P
ผู้ประเมินคนที่ 1 / คนที่ 2	0.47	0.152

ข. ผลการทดสอบปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน (หมายเหตุ: แชตจีพีที x ผู้ประเมิน เป็นกลุ่มควบคุม)

รายการประเมิน	ตัวชี้วัดสถิติ	ค่าประมาณจุด	P
ไปตู้แปลภาษา x ผู้ประเมิน	<i>Estimate</i> _{Baidu:R2}	-0.84	0.511
กูเกิลแปลภาษา x ผู้ประเมิน	<i>Estimate</i> _{Google:R2}	-0.82	0.525

ค. การเปรียบเทียบคุณภาพของเครื่องมือแปลเป็นรายคู่ด้วยค่า odds ratio

คู่เปรียบเทียบ	ค่า odds ratio (OR)	P
แชตจีพีที vs. ไปตู้แปลภาษา	35.01	< .001
แชตจีพีที vs. กูเกิลแปลภาษา	13.52	0.010
ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	0.39	0.122

ยอมรับได้ดังนี้ แชตจีพีที (ร้อยละ 99.96; 95%CI: 99.74-100.00) ไปตู้แปลภาษา (ร้อยละ 99.25; 95%CI: 97.16-100) และกูเกิลแปลภาษา (ร้อยละ 99.11; 95%CI: 96.80-100)

ผลการวิเคราะห์ด้วย Bayesian generalized linear mixed models พบว่า ไม่มีปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมินอย่างมีนัยสำคัญทางสถิติ

($P > 0.05$) ผู้ประเมินไม่ส่งผลกระทบต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญ ($P = 0.203$) (ตารางที่ 3 ก และ ข)

การเปรียบเทียบรายคู่พบว่า แชตจีพีทีมีคะแนนการยอมรับได้สูงกว่าไปตู้แปลภาษาและกูเกิลแปลภาษาอย่างมีนัยสำคัญทางสถิติ ($P < 0.004$ และ $P_p = 0.002$ ตามลำดับ) ขณะที่ไปตู้แปลภาษาและกูเกิลแปลภาษามีคะแนนยอมรับได้ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($P = 0.951$) (ตารางที่ 3 ค)

ตารางที่ 3. ผลของการวิเคราะห์ด้วย Bayesian generalized linear mixed models เพื่อทำนายการยอมรับได้ของผลการแปลระดับคำศัพท์จากภาษาไทยเป็นภาษาจีน

ก. ผลการทดสอบอิทธิพลหลักของผู้ประเมินผลของการแปลระดับคำศัพท์

รายการประเมิน	ค่า odds ratio (OR)	P
ผู้ประเมินคนที่ 1 / คนที่ 2	0.48	0.203

ข. ผลการทดสอบปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน (หมายเหตุ: แชตจีพีที x ผู้ประเมิน เป็นกลุ่มควบคุม)

รายการประเมิน	ตัวชี้วัดสถิติ	ค่าประมาณจุด	P
ไปตู้แปลภาษา x ผู้ประเมิน	<i>Estimate</i> _{Baidu:R2}	-0.73	0.574
กูเกิลแปลภาษา x ผู้ประเมิน	<i>Estimate</i> _{Google:R2}	-1.05	0.418

ค. การเปรียบเทียบคุณภาพของเครื่องมือแปลเป็นรายคู่ด้วยค่า odds ratio

คู่เปรียบเทียบ	ค่า odds ratio (OR)	P
แชตจีพีที vs. ไปตู้แปลภาษา	19.65	0.004
แชตจีพีที vs. กูเกิลแปลภาษา	23.24	0.002
ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	1.18	0.951

ตารางที่ 4. คะแนนคุณภาพการแปลระดับวลีหรือประโยคจากภาษาจีนเป็นภาษาไทยจำแนกตามด้านของคุณภาพและเครื่องมือแปล

ด้านที่ประเมิน	เครื่องมือ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่ามัธยฐาน	ค่าต่ำสุด	ค่าสูงสุด
adequacy	แชตจีพีที	4.95	0.25	5	3	5
	กูเกิลแปลภาษา	4.79	0.67	5	1	5
	ไปตู้แปลภาษา	4.62	0.87	5	1	5
fluency	แชตจีพีที	4.88	0.38	5	3	5
	กูเกิลแปลภาษา	4.59	0.74	5	1	5
	ไปตู้แปลภาษา	4.21	1.10	5	1	5
meaning	แชตจีพีที	4.77	0.61	5	1	5
	กูเกิลแปลภาษา	4.40	0.98	5	1	5
	ไปตู้แปลภาษา	4.01	1.24	5	1	5
severity	แชตจีพีที	4.79	0.66	5	1	5
	กูเกิลแปลภาษา	4.53	0.93	5	1	5
	ไปตู้แปลภาษา	4.15	1.19	5	1	5

คุณภาพของการแปลระดับวลีหรือประโยค

ผลการประเมินคุณภาพของการแปลระดับวลีหรือประโยคโดยรวมระหว่างผู้ประเมิน 2 ท่านพบว่ามีความสอดคล้องกันในระดับปานกลาง ($Kappa = 0.51$) ผลการประเมินคุณภาพในด้าน adequacy และ fluency พบว่ามีความสอดคล้องกันในระดับพอใช้ ($Kappa = 0.39$ และ 0.36 ตามลำดับ) ผลการประเมินคุณภาพในด้าน meaning และ severity พบว่ามีความสอดคล้องกันในระดับดีและปานกลางตามลำดับ ($Kappa = 0.62$ และ 0.56 ตามลำดับ)

การแปลวลีหรือประโยคจากจีนเป็นไทย

ค่าเฉลี่ยของคะแนนคุณภาพการแปลระดับวลีหรือประโยคจากภาษาจีนเป็นภาษาไทยแสดงอยู่ในตารางที่ 4 แชตจีพีทีมีค่าเฉลี่ยสูงสุดและมีค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation: SD) ต่ำสุดเมื่อเทียบกับเครื่องมืออื่นในทุกด้านของการประเมิน (ตารางที่ 4)

ผลของการวิเคราะห์ด้วย ordinal logistic mixed-effects model เพื่อทำนายคุณภาพในด้านต่าง ๆ ของการแปลระดับวลีหรือประโยคจากภาษาจีนเป็นภาษาไทยแสดงอยู่ในตารางที่ 5 เครื่องมือแปลและผู้ประเมินไม่มีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติในทุกด้านของการประเมิน ($P > 0.05$) (ตารางที่ 5 ก) ผู้ประเมินไม่ส่งผลกระทบต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญในด้าน adequacy และ meaning ($P = 0.541$ และ 0.060 ตามลำดับ) แต่ส่งผลกระทบต่อคะแนน

คุณภาพการแปลอย่างมีนัยสำคัญในด้าน fluency และ severity ($P < .001$ และ $< .001$ ตามลำดับ) (ตารางที่ 5 ก)

เครื่องมือแปลภาษาส่งผลกระทบต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญ ($P < .001$) ในทุกด้านของการประเมิน (ตารางที่ 5 ก) ผลการเปรียบเทียบเครื่องมือแปลภาษารายคู่พบว่า แชตจีพีทีมีคุณภาพสูงกว่าไปตู้แปลภาษาและกูเกิลแปลภาษาอย่างมีนัยสำคัญทางสถิติในทุกด้านของการประเมิน ส่วนกูเกิลแปลภาษามีคุณภาพสูงกว่าไปตู้แปลภาษาอย่างมีนัยสำคัญในทุกด้านของการประเมิน (ตารางที่ 5 ข)

การแปลวลีหรือประโยคจากไทยเป็นจีน

ค่าเฉลี่ยของคะแนนคุณภาพการแปลระดับวลีหรือประโยคจากภาษาไทยเป็นภาษาจีนแสดงอยู่ในตารางที่ 6 แชตจีพีทีมีค่าเฉลี่ยสูงสุดและมีค่า SD ต่ำสุดเมื่อเทียบกับเครื่องมืออื่นในทุกด้านของการประเมิน (ตารางที่ 6)

ผลของการวิเคราะห์ด้วย ordinal logistic mixed-effects model เพื่อทำนายคุณภาพในด้านต่าง ๆ ของการแปลระดับวลีหรือประโยคจากภาษาไทยเป็นภาษาจีนแสดงอยู่ในตารางที่ 7 เครื่องมือแปลและผู้ประเมินไม่มีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติในทุกด้านของการประเมิน ($P > 0.05$) (ตารางที่ 7 ก) ผู้ประเมินไม่ส่งผลกระทบต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญในด้าน adequacy และ meaning ($P = 0.566$ และ 0.325 ตามลำดับ) แต่ส่งผลกระทบต่อคะแนน

ตารางที่ 5. ผลของการวิเคราะห์ด้วย ordinal logistic mixed-effects model เพื่อทำนายคุณภาพในด้านต่าง ๆ ของการแปลระดับวลีหรือประโยคจากภาษาจีนเป็นภาษาไทย

ก. ผลการวิเคราะห์อิทธิพลหลัก (เครื่องมือและผู้ประเมิน) และปฏิสัมพันธ์ระหว่างเครื่องมือและผู้ประเมิน

ด้านที่ประเมิน	แหล่งความแปรปรวน	P
adequacy	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	0.541
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.251
fluency	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	< .001
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.144
meaning	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	0.060
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.427
severity	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	< .001
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.102

ข. ผลการเปรียบเทียบเครื่องมือแปลภาษาแบบรายคู่

ด้านที่ประเมิน	คู่เปรียบเทียบ	ค่าประมาณการ	P
adequacy	แชตจีพีที vs. ไปตู้แปลภาษา	2.731	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	1.769	.005
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-0.962	.016
fluency	แชตจีพีที vs. ไปตู้แปลภาษา	2.890	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	1.710	< .001
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-1.170	< .001
meaning	แชตจีพีที vs. ไปตู้แปลภาษา	2.660	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	1.660	< .001
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-1.000	< .001
severity	แชตจีพีที vs. ไปตู้แปลภาษา	1.665	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	0.902	< .001
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-0.763	< .001

คุณภาพการแปลอย่างมีนัยสำคัญในด้าน fluency และ severity ($P < .001$ และ $< .001$ ตามลำดับ) (ตารางที่ 7 ก)

เครื่องมือแปลภาษาส่งผลต่อคะแนนคุณภาพการแปลอย่างมีนัยสำคัญ ($P < .001$) ในทุกด้านของการประเมิน (ตารางที่ 7 ก) ผลการเปรียบเทียบเครื่องมือแปลภาษารายคู่พบว่า แชตจีพีทีที่มีคุณภาพสูงกว่าไปตู้แปลภาษาและกูเกิลแปลภาษาอย่างมีนัยสำคัญทางสถิติในทุกด้านของการประเมิน ส่วนกูเกิลแปลภาษามีคุณภาพสูงกว่าไปตู้แปลภาษาอย่างมีนัยสำคัญในทุกด้านของการประเมิน

ยกเว้นในด้าน fluency ที่กูเกิลแปลภาษาและไปตู้แปลภาษามีคุณภาพไม่แตกต่างกันอย่างมีนัยสำคัญ ($P = 0.424$) (ตารางที่ 7 ข)

ความผิดพลาดในการแปล

เครื่องมือแปลภาษาทั้งสามมีข้อผิดพลาดในการแปลร่วมกัน คือ ข้อผิดพลาดจากการแปลตรงตัวโดยไม่คำนึงถึงบริบท เช่น คำว่า “闪电症状” ซึ่งในบริบททาง

ตารางที่ 6. คะแนนคุณภาพการแปลระดับวลีหรือประโยคจากภาษาไทยเป็นภาษาจีนจำแนกตามด้านของคุณภาพและเครื่องมือแปล

ด้านที่ประเมิน	เครื่องมือ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่ามัธยฐาน	ค่าต่ำสุด	ค่าสูงสุด
adequacy	แซตจีพีที	4.97	0.17	5	4	5
	กูเกิลแปลภาษา	4.80	0.56	5	1	5
	ไปตู้แปลภาษา	4.61	0.79	5	1	5
fluency	แซตจีพีที	4.94	0.25	5	4	5
	กูเกิลแปลภาษา	4.77	0.53	5	1	5
	ไปตู้แปลภาษา	4.70	0.56	5	2	5
meaning	แซตจีพีที	4.98	0.17	5	3	5
	กูเกิลแปลภาษา	4.76	0.61	5	2	5
	ไปตู้แปลภาษา	4.49	0.95	5	1	5
severity	แซตจีพีที	4.96	0.26	5	3	5
	กูเกิลแปลภาษา	4.74	0.65	5	1	5
	ไปตู้แปลภาษา	4.49	0.99	5	1	5

การแพทย์หมายถึง “อาการเห็นแสงวูบวาบ” แต่ “闪电” ในความหมายทั่วไปมีความหมายว่า “ฟ้าผ่า” ข้อความจึงถูกแปลเป็น อาการฟ้าผ่า

ข้อผิดพลาดที่พบร่วมกันระหว่างกูเกิลแปลภาษาและไปตู้แปลภาษา คือ ข้อผิดพลาดจากการแปลทับศัพท์เนื่องจากขาดคลังคำศัพท์เฉพาะ เช่น “拉多吗?” ซึ่งในบริบททางการแพทย์มีความหมายว่า “ถ่ายเยอะไหม” ถูกแปลเป็น “เป็นลาได้ไหม?” ซึ่งเป็นการแปลทับศัพท์ที่ไม่มีความหมายในภาษาไทย นอกจากนี้ ในไปตู้แปลภาษายังพบข้อผิดพลาดในการแปลสลับอวัยวะ เช่น “ปวดบริเวณลิ้นปี่” ถูกแปลเป็น “舌头疼痛” ซึ่งมีความหมายว่า ปวดลิ้น คือแปลผิดจาก “ลิ้นปี่” กลายเป็น “ลิ้น” ในปาก

การอภิปรายและสรุปผล

การเปรียบเทียบคำสำหรับแซตจีพีทีสามคำสั่งที่กำหนดในการวิจัยพบว่า ไม่มีปฏิสัมพันธ์ของคำสั่งที่ใช้และผู้ประเมินในทุกด้านของคุณภาพที่ประเมินและในทั้งสองทิศทางการแปล คำสั่งทั้งสามที่กำหนดไม่มีความแตกต่างอย่างมีนัยสำคัญต่อคุณภาพการแปลจากภาษาจีนเป็นภาษาไทย และภาษาไทยเป็นภาษาจีน แต่พบว่าตัวผู้ประเมินมีผลต่อการประเมินคุณภาพอย่างมีนัยสำคัญทางสถิติในบางด้านของคุณภาพที่ประเมิน อย่างไรก็ตาม เนื่องจากไม่พบความแตกต่างอย่างมีนัยสำคัญระหว่างคำสั่ง

การศึกษาจึงเลือกคำสั่งการแปลที่มีค่าเฉลี่ยของคุณภาพที่สูงที่สุดและ SE ต่ำที่สุดในแต่ละทิศทางการแปล โดยในทิศทางการแปลจากภาษาจีนเป็นภาษาไทย ใช้คำสั่ง “As a health professional who is a native speaker of the target language, translate “.....” into Thai. Provide fluent and adequate translated text.” ส่วนการแปลจากภาษาไทยเป็นภาษาจีนใช้ “translate “.....” into Chinese” เป็นคำสั่งสำหรับแปลภาษาต่อไปในงานวิจัยนี้

การเปรียบเทียบคุณภาพการแปลด้วยเครื่องมือ 3 ชนิดทั้งในระดับคำศัพท์ครอบคลุมทิศทางการแปลทั้งสองทิศทางพบว่า ไม่มีปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมินอย่างมีนัยสำคัญทางสถิติ และไม่พบความแตกต่างของคุณภาพการแปลอย่างมีนัยสำคัญทางสถิติในระหว่างผู้ประเมิน ผลการแปลคำศัพท์โดยแซตจีพีทีมีคะแนนการยอมรับได้สูงสุดซึ่งเหนือกว่ากูเกิลแปลภาษาและไปตู้แปลภาษาอย่างมีนัยสำคัญทางสถิติ สำหรับกูเกิลแปลภาษาและไปตู้แปลภาษานั้นมีคุณภาพการแปลคำศัพท์ไม่แตกต่างอย่างมีนัยสำคัญทางสถิติ

ส่วนการแปลในระดับวลีหรือประโยคพบว่า ไม่มีปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมินอย่างมีนัยสำคัญทางสถิติในทุกมิติของคุณภาพการประเมินและในทุกทิศทางการแปล ผู้ประเมินทั้งสองให้คะแนนคุณภาพต่างกันในทุกมิติของ fluency และ severity ในทุกทิศทางการแปล แต่ไม่พบความแตกต่างระหว่างผู้ประเมินในการ

ตารางที่ 7. ผลของการวิเคราะห์ด้วย ordinal logistic mixed-effects model เพื่อทำนายคุณภาพในด้านต่าง ๆ ของการแปลระดับวลีหรือประโยคจากภาษาไทยเป็นภาษาจีน

ก. ผลการวิเคราะห์อิทธิพลหลัก (เครื่องมือและผู้ประเมิน) และปฏิสัมพันธ์ระหว่างเครื่องมือและผู้ประเมิน

ด้านที่ประเมิน	แหล่งความแปรปรวน	P
adequacy	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	0.566
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.420
fluency	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	< .001
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.852
meaning	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	0.325
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.880
severity	อิทธิพลของเครื่องมือ	< .001
	อิทธิพลของผู้ประเมิน	< .001
	ปฏิสัมพันธ์ระหว่างเครื่องมือแปลภาษาและผู้ประเมิน	0.093

ข. ผลการเปรียบเทียบเครื่องมือแปลภาษาแบบรายคู่

ด้านที่ประเมิน	คู่เปรียบเทียบ	ค่าประมาณการ	P
adequacy	แชตจีพีที vs. ไปตู้แปลภาษา	3.010	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	2.116	.002
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-0.894	.013
fluency	แชตจีพีที vs. ไปตู้แปลภาษา	2.283	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	1.800	.008
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-0.483	0.424
meaning	แชตจีพีที vs. ไปตู้แปลภาษา	4.590	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	3.470	< .001
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-1.110	.002
severity	แชตจีพีที vs. ไปตู้แปลภาษา	3.250	< .001
	แชตจีพีที vs. กูเกิลแปลภาษา	2.200	.002
	ไปตู้แปลภาษา vs. กูเกิลแปลภาษา	-1.050	.007

ประเมินด้าน adequacy และ meaning ซึ่งอาจเป็นไปได้ว่าด้าน adequacy และ meaning เป็นการประเมินความครบถ้วนและความหมายของเนื้อหาจากการแปลเมื่อเทียบกับเคียงกับภาษาต้นฉบับ จึงมีความเป็นรูปธรรมที่ไม่ต้องใช้ดุลพินิจส่วนบุคคล ทำให้เกิดความไม่สอดคล้องระหว่างผู้ประเมินน้อย แต่การประเมินด้าน fluency และ severity เป็นการประเมินความเป็นธรรมชาติของคำแปลที่ได้ ความถูกต้องตามไวยากรณ์ และผลกระทบของความคลาดเคลื่อนในการแปลต่อสุขภาพ ซึ่งต้องอาศัยความเห็น

ส่วนตัวของผู้ประเมินประกอบ จึงทำให้พบความไม่สอดคล้องระหว่างผู้ประเมินได้ง่ายกว่า

ในการแปลระดับวลีหรือประโยค แชตจีพีทีมีคุณภาพดีที่สุดเช่นกันในการแปลภาษาทั้งสองทิศทาง โดยมีค่าเฉลี่ยของคุณภาพสูงสุดและค่า SD ต่ำสุดในหัวข้อประเมินทั้ง 4 ด้าน (adequacy, fluency, meaning และ severity) ซึ่งสะท้อนถึงความถูกต้องและความสม่ำเสมอในการแปล นอกจากนี้ผู้ประเมินให้คะแนนที่ต่ำที่สุดในการประเมินแชตจีพีทีสูงกว่าเครื่องมืออื่นในเกือบทุกด้าน การ

เปรียบเทียบรายคู่พบว่า แชตจีพีทีที่มีคุณภาพดีที่สุดในการประเมินทั้ง 4 ด้าน ภูเก็ตแปลภาษามีคุณภาพทั้ง 4 ด้านสูงกว่าไปตู้แปลภาษาอย่างมีนัยสำคัญในการแปลจากภาษาจีนเป็นภาษาไทย ส่วนทิศทางการแปลจากภาษาไทยเป็นภาษาจีนนั้น ระยะห่างของคุณภาพระหว่างภูเก็ตแปลภาษาและไปตู้แปลภาษาลดลงจนไม่มีความแตกต่างกันอย่างมีนัยสำคัญในด้าน fluency

ผลการวิจัยโดยภาพรวมชี้ให้เห็นชัดเจนว่า แชตจีพีทีที่มีคุณภาพเหนือกว่าทั้งภูเก็ตแปลภาษาและไปตู้แปลภาษาในการแปลระดับคำศัพท์และระดับวลีหรือประโยค โดยเฉพาะในระดับประโยค ซึ่งอธิบายได้ว่า แชตจีพีทีที่มีสถาปัตยกรรมที่เข้าใจบริบทได้ลึกซึ้งกว่าเครื่องมือแปลภาษาแบบ neural machine translation ซึ่งเป็นการแปลแบบดั้งเดิม ทำให้แชตจีพีทีที่สามารถรักษาใจความสำคัญได้สมบูรณ์กว่าแม้ในประโยคที่มีความซับซ้อนสูง หากต้องการใช้เครื่องมือแปลภาษา ผู้ใช้ควรเลือกใช้งานแชตจีพีทีที่เป็นเครื่องมือหลัก อย่างไรก็ตามเมื่อพิจารณาการใช้งานจริง พบประเด็นที่น่าสนใจ คือ เครื่องมือแปลภาษาแบบดั้งเดิมทั้งภูเก็ตแปลภาษาและไปตู้แปลภาษายังคงมีข้อได้เปรียบในด้านความสะดวกและรวดเร็ว เนื่องจากผู้ใช้สามารถนำข้อความเข้าสู่ระบบเพื่อแปลได้ทันทีโดยไม่ต้องอาศัยเทคนิคการสร้างคำสั่ง

เมื่อเปรียบเทียบกับวรรณกรรมก่อนหน้า ผลการวิจัยนี้สอดคล้องกับงานวิจัยในประเมินการแปลภาษาไทยเป็นจีนในบริบททั่วไปของกัลยา ขาวบ้านแพ้ว (21) ที่พบผลการประเมินที่สอดคล้องกันว่า ภูเก็ตแปลภาษามีสัดส่วนคุณภาพในการแปลดีกว่าไปตู้แปลภาษา นอกจากนี้ Zuo และคณะ (12) พบลำดับคุณภาพของทั้ง 3 เครื่องมือเช่นเดียวกัน คือ แชตจีพีทีที่มีประสิทธิภาพสูงสุด รองลงมาคือภูเก็ตแปลภาษา และไปตู้แปลภาษา อย่างไรก็ตาม งานดังกล่าวที่ทำในปี 2024 พบว่า คะแนนความคล้ายคลึงของการแปลจากแชตจีพีทีอยู่ในระดับค่อนข้างต่ำ (12) แต่ในงานวิจัยครั้งนี้คะแนนคุณภาพการแปลในทุกด้านของแชตจีพีทีอยู่ในระดับที่ค่อนข้างสูง แสดงให้เห็นว่า โมเดลมีการพัฒนาให้มีคุณภาพมากยิ่งขึ้น

งานวิจัยในอดีตที่ศึกษาการแปลระหว่างภาษาอังกฤษและภาษาจีนในบริบททางการแพทย์ โดย Chen และคณะ (22) และ Taira และคณะ (23) พบว่าภูเก็ตแปลภาษายังคงมีคะแนนคุณภาพส่วนใหญ่อยู่ในระดับปานกลางถึงค่อนข้างสูง ซึ่งสอดคล้องกับผลจากงานวิจัยในครั้ง

นี้ งานวิจัยของ Khanna และคณะ (24) ประเมินคุณภาพการแปลรายงานทางรังสีวิทยา ซึ่งผู้ประเมินมีความเห็นด้วยและเห็นด้วยอย่างยิ่งในระดับสูงต่อผลการแปลของแชตจีพีที ซึ่งสอดคล้องกับงานวิจัยนี้ที่ผู้ประเมินให้คะแนนคุณภาพของแชตจีพีทีค่อนข้างสูงเช่นกัน

ในกรณีที่ต้องการแปลข้อมูลส่วนบุคคลด้านสุขภาพรวมถึงประวัติการเจ็บป่วยของผู้ป่วยโดยป้อนลงในแอปพลิเคชันหรือเซิร์ฟเวอร์สาธารณะนั้น ต้องทำด้วยความระมัดระวังเพราะมีความเสี่ยงสูงต่อการละเมิดกฎหมายคุ้มครองข้อมูลส่วนบุคคลและการรั่วไหลของข้อมูลทางการแพทย์ ดังนั้น การใช้งานเครื่องมือแปลในกรณีข้างต้นอาจต้องเปลี่ยนไปใช้ระบบที่เป็นเวอร์ชันสำหรับองค์กรหรือระบบที่ทำงานแบบปิดภายในองค์กรเพื่อความปลอดภัย ซึ่งข้อจำกัดที่ตามมาคือ คุณภาพในการแปลและประมวลผลทางภาษาของระบบปิดเหล่านั้น อาจแตกต่างออกไปเมื่อเทียบกับแอปพลิเคชันเวอร์ชันสาธารณะที่ใช้ในงานวิจัยชิ้นนี้

เมื่อพิจารณาผ่านบริบทของการซักประวัติผู้ป่วยซึ่งเป็นขั้นตอนที่มีความเสี่ยงสูงและเกี่ยวข้องโดยตรงกับความปลอดภัยของผู้ป่วย ผู้วิจัยพบข้อสังเกตสำคัญคือ แม้เครื่องมือที่มีคุณภาพสูงอย่างแชตจีพีทีจะมีคะแนนคุณภาพที่ดี แต่ยังพบความคลาดเคลื่อนในการแปลที่อาจส่งผลกระทบต่อสุขภาพของผู้ป่วย คำแปลที่ได้คะแนน severity ระดับ 1 (ทำให้เกิดอันตรายต่อสุขภาพ) เช่น “闪电症状” ที่มีความหมายว่า อาการเห็นแสงวูบวาบ ถูกแปลเป็น “อาการฟ้าผ่า” หรือ “ปวดบริเวณลิ้นปี่” ถูกแปลเป็น “舌头疼痛” ที่มีความหมายว่า อาการเจ็บลิ้น ซึ่งอาจนำไปสู่การตีความที่ผิดพลาดในประเด็นสำคัญทางการแพทย์ได้ คะแนนคุณภาพเฉลี่ยที่สูงโดยรวม อาจไม่ได้เป็นเครื่องยืนยันว่า ผลการแปลมีความปลอดภัยในการนำไปใช้งานจริงโดยปราศจากการตรวจสอบจากมนุษย์

งานวิจัยนี้มุ่งเน้นการทดสอบชุดข้อความที่เกี่ยวข้องกับการซักประวัติเบื้องต้นทั่วไป และการซักประวัติเบื้องต้นในระบบประสาท ระบบทางเดินอาหาร และระบบสืบพันธุ์เท่านั้น ผู้ใช้ควรตระหนักว่า ผลการวิจัยนี้ยังไม่ครอบคลุมคำศัพท์ในระบบอื่นและคำศัพท์เฉพาะทางการแพทย์ขั้นสูง โรคที่มีความซับซ้อนสูง หรือภาวะฉุกเฉินทางการแพทย์ ดังนั้นควรระมัดระวังในการนำผลการศึกษาไปประยุกต์ใช้กับข้อมูลที่มีความซับซ้อน มีความจำเพาะ

เจาะจงทางเทคนิคในทางการแพทย์สูง หรือบริบททาง การแพทย์ที่มีความเสี่ยงสูง

การวิจัยนี้มุ่งเน้นการวิเคราะห์เชิงปริมาณเพื่อ เปรียบเทียบคุณภาพของเครื่องมือแปลภาษาด้วยตัวชี้วัด ทางสถิติเป็นหลัก ไม่มีการวิเคราะห์ถึงสาเหตุและจำแนก ข้อผิดพลาดเชิงคุณภาพของเครื่องมือเหล่านั้น เช่น ข้อผิดพลาดที่เกิดจากความแตกต่างทางวัฒนธรรมและ สำนวนภาษา หรือข้อผิดพลาดที่เกิดจากความ เฉพาะเจาะจงของคำศัพท์ทางการแพทย์ ซึ่งเป็นข้อมูล สำคัญที่ใช้ในการพัฒนาโมเดลต่อไป

การประเมินด้าน severity (ความคลาดเคลื่อนใน การแปลที่ส่งผลกระทบต่อสุขภาพ) ถือเป็นด้านที่สำคัญใน บริบททางคลินิก ซึ่งมีจำนวนผู้ประเมิน 2 ท่าน แม้ผู้วิจัยจะ พยายามลดความแปรปรวนของข้อมูลด้วยหลักการของ Bayesian หรือ mixed-effects model แต่ค่าความสอดคล้อง ของผู้ประเมินอาจไม่เพียงพอสำหรับงานวิจัยด้านสุขภาพ ดิจิทัลที่ต้องการความแม่นยำสูง ทั้งนี้ ข้อจำกัดดังกล่าวเกิด จากความยากในการคัดเลือกผู้เชี่ยวชาญที่ต้องมีความ ชำนาญทางการแพทย์ควบคู่ไปกับทักษะทางภาษาจีนใน ระดับสูง ซึ่งหาได้ยากในทางปฏิบัติ

ข้อเสนอแนะ

เครื่องมือแปลภาษาควรใช้เป็นเพียงตัวช่วย เบื้องต้นในการสื่อสารเท่านั้น หรือควรใช้เพื่อการคัดกรองใน ด้านแรก เนื่องจากผลการประเมินด้าน severity ยังพบ ข้อผิดพลาดในการแปลที่อาจทำให้เกิดอันตรายต่อสุขภาพ ในการใช้งานเครื่องมือเหล่านี้ บุคลากรทางการแพทย์ควร ตรวจสอบประเด็นสำคัญอื่น ๆ เพิ่มเติม เพื่อป้องกัน ความคลาดเคลื่อนที่อาจส่งผลกระทบต่อสุขภาพของผู้รับบริการ นอกจากนี้ เพื่อให้การใช้งานเครื่องมือเหล่านี้มีความ ปลอดภัยและสอดคล้องกับจริยธรรมทางการแพทย์ หาก บุคลากรทางการแพทย์ต้องการนำเครื่องมือเหล่านี้ไป ประยุกต์ใช้ในการทำงานจริง ควรหลีกเลี่ยงการกรอกข้อมูล ที่ระบุตัวตนผู้ป่วย เนื่องจากระบบเหล่านี้ยังไม่สอดคล้องกับ พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 ของ ประเทศไทย

ในอนาคตควรพิจารณาเพิ่มจำนวนผู้ประเมินเป็น 3-5 ท่าน เพื่อลดความแปรปรวนจากดุลยพินิจส่วนบุคคล และกำหนดให้มีผู้เชี่ยวชาญที่เป็นเจ้าของภาษาจีนร่วมเป็น ผู้ประเมินด้วย นอกจากนี้ ควรเพิ่มรายละเอียดในคู่มือการ

ประเมินให้มีความเฉพาะเจาะจงมากขึ้น โดยเฉพาะในมิติ fluency เพื่อลดช่องว่างของการใช้ดุลยพินิจส่วนบุคคลให้ น้อยลง การศึกษาในอนาคต ควรขยายขอบเขตชุดข้อมูล ทดสอบให้ครอบคลุมภาษาพูด แสลง และภาษาถิ่นที่ผู้ป่วย นิยมใช้ในการสื่อสารจริงในงาน เพื่อประเมิน ความสามารถของเครื่องมือแปลภาษาในการตีความ ข้อความที่มีความกำกวมสูง การศึกษาในอนาคตอาจใช้ เสียงจริงในการทดสอบเครื่องมือแปลภาษา เพื่อจำลอง สถานการณ์ให้ใกล้เคียงกับการชกประวัติจริง ซึ่งจะช่วยการ ทดสอบสอดคล้องกับการปฏิบัติงานจริงในคลินิก หรือควร เปรียบเทียบคุณภาพของเครื่องมือในระดับข้อมูลที่ซับซ้อน มากกว่าการชกประวัติเบื้องต้น เพื่อให้ทราบขีดจำกัดของ เครื่องมือในการประมวลผลศัพท์เฉพาะทาง การศึกษาใน อนาคตควรเปรียบเทียบผลลัพธ์ที่กับโมเดลอื่น ๆ เช่น Claude หรือ Gemini เพื่อศึกษาพัฒนาการของเทคโนโลยี การแปลในกลุ่มโมเดลภาษาขนาดใหญ่

กิตติกรรมประกาศ

การวิจัยครั้งนี้สำเร็จได้ด้วยความอนุเคราะห์จากผู้ ประเมินทุกท่าน งานวิจัยนี้ได้รับทุนอุดหนุนการทำ วิทยานิพนธ์ของนักศึกษาระดับบัณฑิตศึกษาจาก เงินกองทุนสนับสนุนการวิจัย นวัตกรรม และการสร้างสรรค์ คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร

การใช้ปัญญาประดิษฐ์ในบทความ

บทความนี้ใช้ Gemini 3 Flash ในการเขียน บทความ โดยลักษณะการใช้งาน คือ ค้นหาวรรณกรรมที่ เกี่ยวข้อง ระดมสมอง ช่วยปรับปรุงและขัดเกลาให้ข้อความ สละสลวย ร่างและย่อบทนำ แปลภาษา ช่วยเขียนและปรับ โคลิดในการวิเคราะห์สถิติ ทั้งนี้ผู้เขียนขอยืนยันว่าได้ ตรวจสอบความถูกต้องของเนื้อหาทั้งหมดที่ AI สร้างขึ้นและ ผู้เขียนขอรับผิดชอบต่อความถูกต้อง ความคิดริเริ่ม และ ความน่าเชื่อถือของเนื้อหาในบทความฉบับสมบูรณ์

เอกสารอ้างอิง

1. Beijing Royal Thai Embassy. The relationship between Thailand and China [online]. 2026 [cited Feb 23, 2026]. Available from: thaiembbeij.org/th/republic-of-china/thai-relations-china/.

2. Department of Business Development. Foreign investment in Thailand reaches 276 billion baht in the first 10 months of 2025, an increase of 115 billion baht or 72% from 2024, with Singapore as the top investor [online]. 2026 [cited May 15, 2026]. Available from: www.dbd.go.th/news/5326112568.
3. Ministry of Commerce. Positive sign! foreign investment in thailand surges to 33 billion baht at the start of 2026, a 46% growth [online]. 2026 [cited May 15, 2026]. Available from: www.moc.go.th/th/gallery/category/detail/id/5/iid/1629.
4. Department of Employment. foreign workers administration office: statistics of permitted foreign workers remaining in the kingdom [online]. 2025 [cited May 15, 2026]. Available from: www.doe.go.th/prd/assets/upload/files/alien_th/be65e4165d12436be6407224633c94bb.pdf.
5. Kedsaboot J. Trends in quantitative and qualitative changes of Chinese students in Thailand from 2019 to 2024. *Journal of Integration Social Sciences and Development* 2025; 8: 350-61.
6. Ministry of Tourism and Sports. International tourist arrivals to Thailand Jan-Dec 2024P [online]. 2025 [cited May 16, 2026]. Available from: www.mots.go.th/news/category/759.
7. Ministry of Tourism and Sports. Tourism statistics 2025 [online]. 2025 [cited May 16, 2026]. Available from: www.mots.go.th/news/category/805.
8. Ministry of Tourism and Sports. International tourist arrivals to thailand 2026 [online]. 2025 [cited May 16, 2026]. Available from: www.mots.go.th/news/category/817.
9. Sittithamrongchai T. Medical tourism budget: what makes Thailand the top destination [online]. 2024 [cited Feb 20, 2026]. Available from: web.parliament.go.th/view/82/News_Detail//1490/TH-TH.
10. Khamrangsi M, Piarak S, Luechai T, Sirirojphon K, Chaiyawong C. How to do effectively history taking: a case study of gynecology system. *Journal of Nurses Association of Thailand Northern Office* 2019 ; 25: 71-9.
11. Google LLC. Google Translate [online]. 2025 [cited Oct 31, 2025]. Available from: translate.google.com.
12. Zuo Y, Ching GS, Khotsing R. The application of ChatGPT in literary translation: a case study from Thai to Chinese. In: *LTEC 2024. Proceedings of the 12th International Workshop on Learning Technology for Education Challenges*; 2024 Jul 29-Aug 1; Kaohsiung, Taiwan. Cham: Springer Nature Switzerland; 2024. p. 332-48.
13. Baidu Inc. Baidu Translate [online]. 2025 [cited Oct 31, 2025]. Available from: fanyi.baidu.com.
14. NetEase Youdao. Youdao Translate [online]. 2025 [cited Oct 31, 2025]. Available from: fanyi.youdao.com.
15. OpenAI. ChatGPT [online]. 2025 [cited Oct 31, 2025]. Available from: chatgpt.com.
16. Wang H, Wu H, He Z, Huang L, Church KW. Progress in machine translation. *Engineering*. 2022; 18:143-53.
17. Zhang J, Zong C. Neural machine translation: challenges, progress and future. *Sci China Technol Sci*. 2020; 63: 2028-50.
18. Somabut A. What is generative AI: perspectives on teaching and learning in higher education [online]. 2025 [cited Jan 10, 2026]. Available from: ltic.kku.ac.th/what-is-generative-ai/.
19. Dzuali F, Seiger K, Novoa R, Aleshin M, Teng J, Lester J, et al. ChatGPT may improve access to language-concordant care for patients with non-English language preferences. *JMIR Med Educ*. 2024; 10: e51435.
20. Thammasat Business School. Prompt engineering: a new skill in the age of generative AI [online]. 2024 [cited Dec 13, 2025]. Available from: conc.tbs.tu.ac.th/index.php/learning/detailprogram/31?srsId=AfmBOoqGtYs50YqrTFYZxgJ_EpAOqJq52n74GE585P2hlbvHTr6MTyEa.

21. Khaowbanpaew K. Problem analysis and quality assessment of translation from Thai to Chinese with Baidu Translate and Google Translate translation tools. *Journal of Liberal Arts, Maejo University* 2022; 10: 109-33.
22. Chen X, Acosta S, Barry AE. Evaluating the accuracy of Google Translate for diabetes education material. *JMIR Diabetes*. 2016; 1: e5848.
23. 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.
24. Khanna P, Dhillon G, Buddhavarapu V, Verma R, Kashyap R, Grewal H. Artificial intelligence in multilingual interpretation and radiology assessment for clinical language evaluation (AI-MIRACLE). *J Pers Med*. 2024; 14: 923.
25. Khanna RR, Karliner LS, Eck M, Vittinghoff E, Koenig CJ, Fang MC. Performance of an online translation tool when applied to patient educational material. *J Hosp Med*. 2011; 6: 519-25.
26. Injai P. Handbook of three languages patient history taking (Thai, English, and Chinese). Chonburi: Division of Research, Academic Services and International Relations, Faculty of Medicine, Burapha University; 2021.
27. Costa-Jussà MR, Farrús M, Serrano J. Machine translation in medicine: a quality analysis of statistical machine translation in the medical domain. Paper presented at: Advanced Research in Scientific Areas; 2012 Dec 3-7; Slovakia.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159-74.