אוניברסיטת בר־אילן  
הפקולטה למדעי הרוח  
המחלקה לתרגום וחקר התרגום

**שימוש בתרגום מכונה בסביבות רפואיות**

סמינריון מאמרים

שם הסטודנטית: נעמה צ'יטיז  
מס' תעודת זהות: 304965700  
שם הקורס: תרגום במאה ה־21   
מס' הקורס: 50972-01   
שם המרצה: ד"ר עמרי אשר

**תוכן עניינים**

|  |  |
| --- | --- |
| מבוא | 3 |
| שימוש בתרגום מכונה במסגרות רפואיות | 4 |
| הערכה של Google Translate עבור תרגום הוראות חדר מיון | 6 |
| הערכת תרגום מכונה בעל פה עבור שיחות פשוטות במסגרת קלינית | 9 |
| שימוש אמין ובטיחותי בתרגום מכונה במסגרות רפואיות | 11 |
| פיתוח תרגום מכונה סווהילית קונגולזית לצורך סיוע הומניטרי במגפת הקורונה | 17 |
| דיון | 22 |
| סיכום | 27 |
| ביבליוגרפיה | 28 |

**מבוא**

הצורך בשירותי בריאות המותאמים לשפות רבות גדל ברחבי העולם. עם הגלובליזציה הגוברת, יש צורך גדול יותר בכלים המשרתים את צורכי השפה המגוונים של תיירים, מהגרים, פליטים ואנשים שגורשו ממולדתם. העשורים האחרונים הביאו לחסמים מופחתים על תנועה של אנשים, ציוד ושירותים, לזמינות רבה יותר של מידע מקוון ולמאמץ משותף של ממשלות ושל המגזר הפרטי כאחד למשוך הכנסות באמצעות מערכות הבריאות שלהם (Lunt et al., 2011).

תהליכי הגלובליזציה וההגירה הובילו לכך שישנן קבוצות גדולות של אנשים שאינם דוברים את השפה הרשמית של המדינה שבה הם מתגוררים, או שאינם דוברים אותה באופן שוטף. ההשלכות של עניין זה בהקשר הרפואי הן שמטופלים עם מחסומי שפה, שאינם דוברים את שפת הצוות הרפואי המטפל בהם, מתמודדים עם קשיים בגישה למשאבי חינוך לבריאות ולקבלת שירותי בריאות. קשיים אלו מובילים הן לעלויות גבוהות יותר בתחום הבריאות והן לפערים בריאותיים גדולים יותר, כמו ביצוע פחות בדיקות בריאותיות שגרתיות ובדיקות רפואה מונעת (Dew et al., 2018).

השימוש בכלי תרגום מכונה הולך וגובר במסגרות ובסביבות שונות, וביניהן בתחום הבריאות והטיפול. כלי תרגום מכונה משתפרים באיכותם באופן תמידי, נוחים לשימוש, זמינים באופן נרחב, ונמצאים בשימוש במגוון הקשרים. אולם, השימוש בתרגום מכונה לתקשורת בהקשר של בריאות טומן בחובו אתגרים וחששות, ביניהם שמירה על דיוק ושטף. תרגום שגוי של מידע רפואי עלול להיות בעל השלכות קשות ולסכן את בריאות המטופל (Dew et al., 2018). השאלה אם תרגום מכונה מתאים ובטוח לשימוש במסגרות רפואיות נחקרה בעבר וממשיכה להיחקר עם התקדמות הטכנולוגיה, ובעבודה זו יוצגו ארבעה מחקרים בתחום הנוגעים בשאלה זו מזוויות שונות.

**שימוש בתרגום מכונה במסגרות רפואיות**

נושא העבודה הוא שימוש בתרגום מכונה בכתב ובעל פה במסגרות רפואיות שונות. הסביבות וההקשרים המופיעים במאמרים כוללים תרגום של טקסטים רפואיים במחלקות שונות בבתי חולים, ומתן הנחיות ומידע חיוני בעת מגפת הקורונה. נבחרו ארבעת המאמרים הבאים:

מאמר 1: הערכה של Google Translate עבור תרגום הוראות חדר מיון

Taira, B. R., Kreger, V., Orue, A., & Diamond, L. C. (2021). A pragmatic assessment of google translate for emergency department instructions. *Journal of General Internal Medicine*, *36*(11), 3361-3365.

מאמר 2: הערכת תרגום מכונה בעל פה עבור שיחות פשוטות במסגרת קלינית

Lee, W., Khoong, E. C., Zeng, B., Rios-Fetchko, F., Ma, Y., Liu, K., & Fernandez, A. (2023). Evaluation of Commercially Available Machine Interpretation Applications for Simple Clinical Communication. *Journal of General Internal Medicine*, 1-7.

מאמר 3: שימוש אמין ובטיחותי בתרגום מכונה במסגרות רפואיות

Mehandru, N., Robertson, S., & Salehi, N. (2022, June). Reliable and safe use of machine translation in medical settings. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2016-2025).

מאמר 4: פיתוח תרגום מכונה סווהילית קונגולזית לצורך סיוע הומניטרי במגפת הקורונה

Öktem, A., DeLuca, E., Bashizi, R., Paquin, E., & Tang, G. (2021). Congolese swahili machine translation for humanitarian response. *arXiv preprint arXiv:2103.10734*.

המשותף לכל המאמרים הוא הדיון בהיבטים מעשיים של שימוש בתרגום מכונה בסביבה רפואית. המאמר הראשון הוא מחקר כמותי העוסק בהערכת הדיוק של Google Translate לתרגום מכתבי שחרור מחדר מיון עבור חולים בעלי שליטה מוגבלת באנגלית. גם המאמר השני הוא מחקר כמותי, והוא בודק את איכות הביצועים של שלושה יישומי תרגום מסחריים לצורך תרגום בעל פה של שיחות בין אנשי צוות רפואי ומטופלים. המאמר השלישי הוא מחקר איכותני העוסק בפתרונות בהם נוקטים קלינאים כאשר הם מטפלים בחולים בעלי שליטה מוגבלת באנגלית, ומסיק מהן מסקנות לגבי דרכי שימוש בטוחות ואמינות יותר בתרגום מכונה בסביבה רפואית. במאמר הרביעי בוצע מחקר כמותי בתחום הבלשנות החישובית, והוא מתאר את מאמציהם של מתרגמים מארגון Translators Without Borders לפתח תרגום מכונה בין השפות סווהילית קונגולזית וצרפתית לצורך מתן מענה הומניטרי בזמן מגפת הקורונה.

תחילה יוצגו כל אחד מהמאמרים בהרחבה, ולאחר מכן יוצגו נקודות מבט שונות על סוגיות מרכזיות העולות מהם. הסוגיה המרכזית הראשונה שתוצג בדיון עוסקת בגישה המקדימה לשימוש בתרגום מכונה – שימוש מושכל בתרגום מכונה, תוך התערבות והתאמת הטקסט לפי צרכי המשתמש על מנת להפיק תרגומים איכותיים, לעומת גישה המעריכה את איכות התרגום מכונה ללא כל התערבות והתאמה מצד המשתמש. הסוגיה השנייה תעסוק בהערכת איכות תרגום המכונה לאור תיאוריית הסקופוס, אם מטרת השימוש בתרגום מכונה בסביבות רפואיות היא להשיג תרגומים מדויקים או תרגומים טובים דיים. הצגת ארבעת המאמרים יחד ודיון בסוגיות המשותפות העולות מהם תוכל להוביל לתובנות לגבי אופן השימוש בתרגום מכונה בסביבות רפואיות, על ההיבטים שיש לשים עליהם דגש בהקשר זה, ועל שינויים והתאמות שניתן לעשות על מנת להשתמש בכלי תרגום מכונה באופן נרחב יותר בעתיד.

**מאמר 1: הערכה של Google Translate עבור תרגום הוראות חדר מיון**

תהליך השחרור מחדר המיון הוא נקודה קריטית במיוחד במונחים של תקשורת בין מטפלים למטופלים. הוראות שחרור כתובות מכילות מידע חשוב על האבחנה, תוכנית הטיפול והמעקב של המטופל. לחולים בעלי שליטה מוגבלת באנגלית (Limited English Proficiency - LEP) המשתחררים מחדר מיון יש קשיים גדולים יותר בהבנת סוגי התורים שעליהם לקבוע בקהילה וסוגי התרופות שעליהם ליטול, והם סובלים משיעור גבוה יותר של טעויות תרופתיות וביקורים חוזרים לא מתוכננים למיון.

בעוד שלרוב בתי החולים בארה"ב יש גישה לסיוע בשפה מדוברת באמצעות מתורגמנים טלפוניים, קיים פער ביכולת לסיוע בתרגום כתוב. בבתי חולים רבים יש דפי הסבר למטופלים עם אבחנות ספציפיות, כגון "זיהום בדרכי הנשימה העליונות", הכתובים מראש במגוון שפות ואנשי צוות רפואיים יכולים להשתמש בהם בקלות כדי לספק חומרים כתובים בשפה המועדפת על המטופל. עם זאת, האתגר עולה כאשר איש הצוות רפואי חייב להעביר הנחיות ספציפיות למטופל, כמו "בוא למרפאת עיניים בשעה 8:00 בבוקר ביום חמישי והבא את הרשומות שלך מבית החולים".

לעתים קרובות, אין מנגנון מוגדר לבקשת תרגומים כתובים במסגרת האקוטית. בעוד שהתגובה האופטימלית במצב זה היא לכתוב את הוראות השחרור של המטופל באנגלית ולתרגם את ההוראות בעל פה למטופל באמצעות מתורגמן מוסמך, מטפלים רבים משתמשים בתרגום מכונה לצורך יעילות. Google Translate (GT) היא אופציה לתרגום כתוב שצוברת יותר ויותר פופולריות, ובחלק ממערכות בתי החולים הפכה למקור הרצוי של תרגומים כתובים, במיוחד עבור הוראות שחרור ספציפיות למטופלים הניתנות לחולים בעלי שליטה מוגבלת באנגלית.

המטרה העיקרית של מחקר זה היא לבצע הערכה פרגמטית של הדיוק של GT עבור התרגום הכתוב של הוראות שחרור נפוצות מחדר מיון שניתנו למטופלים בכל אחת מהשפות הנפוצות ביותר המדוברות על ידי חולי LEP בארה"ב כפי שהוערכו על ידי חברי קהילה דו־לשונית. המטרה המשנית היא להשוות את הביצועים של GT בין שפות שונות.

**שיטת המחקר**

הליך המחקר: עורכי המחקר בחרו הנחיות נפוצות שנכתבו בעת שחרור מטופלים מביקור במיון המעבירות מידע קריטי על הטיפול או על תוכנית המעקב. הם בנו רשימה של הצהרות המשקפות הנחיות כתובות שמקבלים מטופלים במכתבי שחרור מחדר מיון. ההצהרות נבדקו לאחר מכן על ידי קבוצה של אנשי צוות רפואי מחדר מיון (Medical Doctor, Registered Nurse & Nurse Practitioner) שאינם מעורבים במחקר. הקבוצה התבקשה להגיב, ועל סמך התגובות נבחרה קבוצה אחרונה של 20 הוראות שחרור מחדר מיון. נבדקו חמש השפות המדוברות ביותר במחוז לוס אנג'לס לצורך המחקר – ספרדית, סינית (כולל מנדרינית וקנטונזית), טאגאלוג (כולל פיליפינית), וייטנאמית וקוריאנית. ארמנית ופרסית נפוצות מאוד בחדר המיון שבו עובדים חלק מהחוקרים והם הוסיפו שפות אלה למחקר – לא רק בגלל התועלת הישירה של הנתונים בסביבה שלהם, אלא גם כדי להשוות את הדיוק של GT עבור שפות בעלות תפוצה נמוכה יותר. כל אחת מ־20 הצהרות הוראות השחרור תורגמו לאחר מכן באמצעות GT לכל שבע שפות היעד.

משתתפי המחקר: אותרו דוברי שפת אם מתנדבים של כל אחת משפות היעד. מתנדבים נכללו אם הם דוברי שפת אם של אחת משפות היעד, דוברי אנגלית שוטפת ויכולים לקרוא את שתי השפות. כדי להבטיח הערכה פרגמטית של הוראות אלו בקרב חברי הקהילה, לא נכללו משתתפים שעבדו בכל תפקיד שהוא של שירותי הבריאות או שהיו מתורגמנים או בלשנים מקצועיים. התקבל אישור IRB לפני תחילת המחקר.

מדדים: הדמוגרפיה הבסיסית של המשתתפים כללה מגדר, שנים בארה"ב, יכולת דיווח עצמי להבין אנגלית ויכולת דיווח עצמית להבין את שפת היעד. בנוסף, התבקשו המתנדבים למלא סולם תרבות בן 4 שאלות.

מדדי תוצאה: המשתתפים קיבלו דף עבודה עם כל אחת מההצהרות שתורגמו ב־GT לשפת האם שלהם והתבקשו להסביר מילולית לחבר צוות המחקר את המשמעות של כל אחת מההצהרות באנגלית. השאלה העיקרית הייתה האם נשמרה כוונת ההצהרה (כן/לא). לאחר מכן, המתנדבים הדו־לשוניים העריכו כל הצהרה לפי הנחיות סטנדרטיות ברובריקת ניקוד של תרגום מכונה. רובריקה זו מכילה סולם ליקרט בן 5 נקודות עבור שטף, התאמה, משמעות וחוּמרה והיא סטנדרטית לדירוג תרגום מכונה. המתנדבים נתנו את הדירוג שלהם לשטף, אדקווטיות ומשמעות וחבר מצוות המחקר בחר את החומרה הקלינית על סמך ההסבר שנתן המתנדב (כלומר, אם ישנה שגיאה בתרגום, האם היא עלולה לסכן את המטופל ובאיזו מידה).

ניתוח נתונים: החוקרים השתמשו בסטטיסטיקה תיאורית כדי לחשב את הדיוק של תרגומי GT של הוראות שחרור מחדר מיון על ידי חישוב שיעור ההיגדים שבהם נשמרה המשמעות מתוך סך ההצהרות. חושבו רווחי סמך של 95%, וחושב שיעור הדיוק עבור כל שפה בנפרד. דווחו הציונים הממוצעים עבור כל אחת מקטגוריות הרובריקה (שטף, אדקווטיות, משמעות וחומרה קלינית).

**תוצאות**

גויסו עשרים מתנדבים ומתנדבות וכל אחד מהם העריך עשרים הוראות שחרור, כך שסך הכול נבדקו 400 הוראות שחרור. המתנדבים היו דוברי ספרדית (5), ארמנית (2), סינית (3), טאגלוג (4), וייטנאמית (2), קוריאנית (2) ופרסית (2). כולם דיווחו בעצמם שהם דוברי אנגלית ברמה טובה (4/20) או טובה מאוד (16/20) ושהם דוברים את שפת היעד ברמה טובה מאוד (18/20) או טובה (2/20). הציונים הממוצעים למידת השטף, אדקווטיות, משמעות וחומרה קלינית היו גבוהים, ונעו בין 4.2 ל-4.4 בסולם ליקרט של 5 נקודות אך השתנו לפי שפה. בסך הכול, GT העבירה במדויק את המשמעות של 330 מתוך 400 (82.5%) הוראות שנבדקו, אך רמת הדיוק השתנתה לפי שפה. הציונים שהתקבלו עבור דיוק בכל אחת מהשפות הם: ספרדית 94%, ארמנית 55%, סינית 81.7%, טאגלוג 90%, ויאטנמית 77.5%, קוריאנית 82.5%, פרסית 67.5%. חלק משגיאות התרגום שדווחו על ידי המתנדבים הפכו את תרגומי GT ללא קוהרנטיים.

**דיון**

ככל שהפרקטיקה של שימוש ב־GT לתקשורת רפואית הופכת נפוצה יותר, חיוני להבין את הדיוק והמגבלות שלה בסביבה הרפואית. במחקר זה, שיעורי הדיוק שנמצאו כפי שהוערכו על ידי מתנדבים מהקהילה עמדו על 6% אי־דיוקים בספרדית ו־18% אי־דיוקים בסינית, אחוזים כמעט זהים לרמות הדיוק של מתרגמים מקצועיים. מידע זה חשוב לעבודה עתידית בתחום מכיוון שההבדל בין תפיסת המטופל של תרגומי מכונה לעומת מתרגמים מקצועיים הינו שאלה שממשיכים לדון בה במחקרים. למרות שעורכי המחקר מצאו שהדיוק הכולל של GT טוב יותר ממה שדווח בעבר, רמות הדיוק השתנו משפה לשפה. באופן מדאיג, ארמנית ופרסית היו בעלי דיוק של 55% ו־67.5% בהתאמה.

מעבר לשונות בשיעורי הדיוק, נמצאו גם כמה בעיות הקשורות לשימוש ב־GT שייתכן כי רופאים עם ידע מוגבל בשפות היעד אינם מודעים להן. לדוגמה, כשהחוקרים יצרו לראשונה את גיליונות העבודה של GT בפרסית, הם גילו שהכיווניות של השפה הכתובה לא נלקחה בחשבון על ידי התוכנה, כלומר שפרסית נכתבת מימין לשמאל. כאשר דף העבודה של GT בפרסית הוצג למתנדב הראשון, הוא היה בלתי קריא מכיוון שהטקסט הועבר משמאל לימין על ידי התוכנה. אם אלו היו הוראות שחרור אמיתיות, הן לא היו קריאות למטופל. מתנדבים גם העלו בעיות פוטנציאליות עם מערכות כתיבה סיניות מסורתיות לעומת מודרניות ובעיות שעלולות לעלות בניבים שונים של פרסית. קלינאי שאינו מודע לבעיות הללו ומשתמש ב־GT באחת משפות אלה עלול לגרום נזק למטופל.

ההשלכה החשובה של מחקר זה היא שלמרות הדיווחים האחרונים על שיפור ברמת הדיוק וההצעה שניתן להשתמש ב־GT בסביבה קלינית, נמצא שהדיוק של GT משתנה באופן מהותי לפי שפה ועדיין אינו כלי אמין לשימוש בקליניקה. גם בשפות שבהן הדיוק גבוה, עדיין קיים פוטנציאל לאי־דיוקים חשובים ופוטנציאל לפגיעה בחולה. השיטה הטובה ביותר נותרה להשתמש בהוראות שחרור כתובות מראש ומתורגמות באופן מקצועי בשפת האם של המטופל לקבלת מידע כללי על אבחנה כאשר דפי מידע כאלה זמינים בתיעוד הבריאות האלקטרוני. להנחיות ספציפיות למטופל, על הרופאים למסור למטופל עותק של הוראות השחרור שלהם באנגלית ולהשתמש במתורגמן כדי לתרגם את ההוראות בעל פה למטופל. בזמן שהמתורגמן נמצא על הקו, יש לבקש מהמטופל שיסביר את ההוראות שניתנו לו (Teach-back) כדי לוודא שהוא הבין את המידע.

**הגבלות**

הגבלות המחקר הן בכך שייתכן שהוא מעריך יתר על המידה את שיעורי הדיוק, מכיוון שהמשתתפים חיו בארה"ב במשך תקופות זמן ארוכות והיו מעורים בתרבות ארה"ב. ייתכן שאינו מייצג את ההבנה של מהגרים חדשים שמולם עומד המחסום הנוסף של חוסר היכרות עם מערכת הבריאות האמריקאית. כל המתנדבים ידעו קרוא וכתוב הן באנגלית הן בשפת היעד ולא בוצעה הערכה רשמית של אוריינות בריאות (health literacy). הדבר עלול גם לגרום להערכת יתר של רמות הדיוק שידווחו על ידי משתתפים בעלי אוריינות בריאות מוגבלת. המשתתפים היו דו־לשוניים וייתכן כי יכולות השפה שלהם לא מייצגות במדויק את ההבנה של מטופלים שהם חד־לשוניים בשפה שאינה אנגלית. לבסוף, GT משתמשת באלגוריתם של בינה מלאכותית שתמיד משתנה וייתכן שנעשו שיפורים נוספים מאז ביצוע המחקר.

**מסקנות**

שיעורי הדיוק של תרגומים על ידי GT עבור הוראות שחרור מחדר מיון משתנים בהתאם לשפה. למרות שהעתיד של תרגום כתוב בבתי חולים הוא ככל הנראה תרגום מכונה, GT אינו מתאים לשימוש כעת במחלקה לרפואה דחופה.

**מאמר 2: הערכת תרגום מכונה בעל פה עבור שיחות פשוטות במסגרת קלינית**

בעוד שמתורגמנות רפואית הינה כלי הכרחי לתקשורת עם מטופלים שאינם דוברים את שפת הצוות הרפואי המטפל בהם, למרבה הצער לרוב אין זה אפשרי או מעשי להשתמש במשאב זה במסגרות קליניות מסוימות, ולכן המשאב אינו מנוצל עד תום. השימוש בשירותי מתורגמנות רפואית פורמליים מאתגר למדי עקב שלל סיבות, ביניהן שטף העבודה העמוס, שמירה על סביבה סטרילית, רמות שונות של הכרה של המטופלים וחילופי שיחות ספורדיים וקצרים. כתוצאה מכך, צוותים רפואיים עלולים לוותר על שימוש במתורגמנים רפואיים ובמקום זאת להסתמך על תקשורת בלתי מילולית, המהווה אתגר למתן טיפול בטוח ואיכותי.

באמצעות תרגום מכונה (Machine Translation - MT) ניתן לגשר על פערים בשפה, הן בכתב הן בדיבור. ישנם יישומי MT חינמיים ופשוטים לתפעול, מאפיינים ההופכים אותם למשאב פרגמטי עבור צוותים רפואיים. עם זאת, שימוש בתרגום מכונה במסגרת רפואית מעלה חששות בטיחותיים מכיוון שתרגום לא מדויק עלול לפגוע באיכות הטיפול.

תרגום מכונה בעל פה מורכב יותר מתרגום מכונה בכתב. נחוצים זיהוי נכון של הדברים הנאמרים, תמלול הדיבור לצורה כתובה ולבסוף סינתזת שפה (יצירת דיבור) כדי ש־MT יפעל כמתורגמן דו־כיווני. כדי לקבוע אם תרגום מכונה בעל פה יכול להיות שימושי לשיחות דו־כיווניות קצרות בעלות סיכון נמוך, בוצע מחקר אי־נחיתות כדי להשוות בין רמות הדיוק והבטיחות של שלושה יישומי MT זמינים מסחרית לבין אלו של שימוש במתורגמנים מקצועיים. התרגומים שנבדקו במחקר הם בין אנגלית לספרדית ובין אנגלית למנדרינית, שתי השפות הנפוצות ביותר בארצות הברית שאינן אנגלית. היישומים שנבדקו הם Microsoft Translator, Google Translate, Apple iTranslate. יישומים אלה זמינים ללא עלות במכשירים שונים ומערכות הפעלה שונות, וכולם משתמשים באלגוריתם למידת מכונה המבוסס על רשתות עצביות מלאכותיות שיכולות להשתפר עם צבירה של נתונים נוספים (Neural Machine Translation).

**שיטת המחקר**

עורכי המחקר כתבו 105 ביטויים באנגלית שנאמרים בדרך כלל בין מטפל למטופל ו־105 ביטויים בספרדית ובמנדרינית שנאמרים בדרך כלל בין מטופל למטפל, המדמים שיחה בין אנשי צוות דוברי אנגלית למטופלים שאינם דוברי אנגלית שפת אם. כל ביטוי כלל אחד עד שלושה משפטים בשפה "רגילה", נטולת סלנג או ביטויים, כגון "האם אתה יכול בבקשה להצביע היכן שכואב?". הביטויים הוקלטו והושמעו לשלושת יישומי ה־MT, והפלט הקולי המתורגם הוקלט. אותם ביטויים מוקלטים הושמעו גם למתורגמנים אנושיים, שתרגמו אותם בעל פה והוקלטו גם הם. ההקלטות של פלט ה־MT ושל המתורגמנים הושמעו לשישה מתנדבים דו־לשוניים שהעריכו את טיב התרגומים ביחס למקור. המתנדבים היו אנשי צוות רפואי כמו גם מתנדבים שאינם מהתחום הרפואי. הם התבקשו להעריך בסולם ליקרט מדדים של דיוק (אובדן מידע בתרגום), שטף (דקדוק), משמעות וסיכון קליני (האם שינוי המשמעות בתרגום עלול לגרום לסכנה למטופל).

**תוצאות**

עבור אנגלית לספרדית, שיעור הביטויים המתורגמים ב־MT שקיבלו ציון מקובל נע בין 0.68 ל־0.84, בעוד שבאנגלית למנדרינית, הטווח היה בין 0.62 ל־0.76. גם תרגום MT מספרדית / מנדרינית לאנגלית היו ציונים מקובלים נמוכים (0.36 עד 0.41). מתורגמנות מאנגלית לספרדית באמצעות Google Translate התקרבה לרף האי־נחיתות, אך אף אחד מהיישומים לא עמד בסף זה.

**דיון**

האיכות הכוללת של שלושת יישומי ה־MT בתרגום בעל פה לשימוש בתקשורת קלינית דו־כיוונית נמצאה ירודה, אפילו בשיחות פשוטות בעלות סיכון נמוך. יישומי ה־MT תרגמו טוב יותר כאשר התרגום היה מאנגלית למנדרינית / ספרדית מאשר להיפך. נמצאו מספר סוגים של שגיאות נפוצות, ביניהן ניתוח תחבירי שגוי וקושי בהבחנה בין הצהרות לשאלות. למשל, יישומי ה­־MT התקשו להבדיל בין "por que?" (למה) ו"porque" (בגלל), בעיה שמתורגמן אנושי לא צריך להתמודד איתה מכיוון שהוא מבחין בין השניים בזכות האינטונציה וההקשר. גם קיצורים רפואיים נפוצים הציבו לפעמים אתגרים, ללא קשר לכיוון התרגום. לדוגמה, הקיצור I.V. פוענח כ־Ivy ותורגם בהתאם.אי־רהיטות, כגון מילות מילוי, גמגום או הפסקות, גם עשויה להשפיע על מתורגמנות MT. דוגמאות למילות מילוי כוללות "אממ", "טוב" ו"you know", שמתורגמנים מקצועיים מתעלמים מהם, אך יישומי MT עשויים לשלב בתרגום שלהם או להפסיק את התרגום לפני השלמת המשפט.

למרות שמחקר זה השווה בין מתורגמנות MT לזו של מתורגמנות רפואית מקצועית, החוקרים מציינים כי הם מודעים לכך שהחלופה הנפוצה ביותר בשיחות פשוטות וקצרות עם סיכון נמוך היא למעשה ללא מתורגמנות כלל. אף על פי כן, על פי החוקרים ממצאי המחקר אינם תומכים כיום בהמלצה לשימוש בתרגום מכונה בעל פה במסגרות קליניות. במקום זאת, הם מעודדים רופאים להשתמש במתורגמנות מקצועית ולהיעזר בחומרה שתאפשר תרגום מרחוק (טלפונים, רמקולים ושיחות וידאו) בכל הסביבות הרפואיות, לפחות עד שתרגום המכונה ישתפר באופן משמעותי עבור תקשורת דו־כיוונית.

עורכי המחקר ציינו את מגבלותיו: הם לא בדקו אם תרגומי מכונה היו מובנים למטופלים. בדיקת ההבנה לוקחת בחשבון את העובדה שהנמענים עשויים להיות מסוגלים להסיק את התוכן המקורי גם אם התרגום לוקה בחסר בדיוק המילוני, הדקדוקי, הסגנוני או הבהירות. כדי לבחון את הבנת המטופלים את התרגום, אפשר לבקש מהם לבצע פעולה ספציפית בעקבות תרגום בעל פה של הוראה. כמו כן, בעולם האמיתי, אדם המשתמש ביישומי MT בדרך כלל יבחין בבעיות העולות תוך כדי פעולה ויתקן אותן, למשל אם היישום יפסיק לתמלל באמצע המשפט הוא ישים לב לכך באמצעות הרמזים החזותיים שמספקות האפליקציות ויחזור על המשפט.

לסיכום, שלושה יישומי MT נפוצים נמצאו כבעלי איכות נחותה בתרגום בעל פה של תקשורת דו־כיוונית בין אנגלית-ספרדית ואנגלית-מנדרינית, אפילו בשיחות פשוטות וקצרות, בהשוואה למתורגמן רפואי מקצועי. עורכי המחקר ממליצים שעד שהאיכות של מתורגמנות MT תשתפר משמעותית, על הרופאים להבטיח טיפול בטוח, יעיל והוגן על ידי עבודה עם מתורגמנים רפואיים מקצועיים במידת האפשר.

**מאמר 3: שימוש אמין ובטיחותי בתרגום מכונה במסגרות רפואיות**

הסיכונים והנזקים של טכנולוגיות תרגום מכונה עלולים לפגוע באופן לא פרופורציונלי באנשים פגיעים התלויים בה לצורך גישה לשירותי בריאות, תעסוקה ותמיכה חברתית. אנשים בדרך כלל משתמשים ב־MT כי אינם יודעים את שפת המקור או את שפת היעד, מה שמקשה מאוד על הערכה בפועל של איכות ותקינות התרגום. במאמר זה החוקרות בוחנות כיצד מערכות MT עלולות להוביל לקצרים בתקשורת ולמידע שגוי בתרחישים בעלי סיכון גבוה – סביבה רפואית. הן בדקו את השימוש בכלי תרגום מכונה על ידי קלינאים (רופאים, אחיות ואנשי צוות רפואיים) ואת ההתמודדות של הקלינאים באופן כללי לנוכח מחסומי שפה עם מטופלים בעלי שליטה מוגבלת באנגלית. על סמך הבעיות שהעלו הקלינאים והפתרונות שבהם נקטו, הן מציעות דרכים לשיפור תרגום מכונה כדי שיהיה בטוח לשימוש במסגרת רפואית.

**שיטות**

מטרת המחקר הייתה לבדוק כיצד מתמודדים קלינאים לנוכח מחסומי שפה מול מטופלים. נערכו ראיונות עומק מובנים למחצה עם 20 קלינאים בשבע התמחויות שונות. הקלינאים הגיעו ממגוון מסגרות ומוסדות – פרקטיקה פרטית, בתי חולים מחוזיים, בתי חולים קהילתיים ומוסדות אקדמיים. החוקרות גייסו את המרואיינים באמצעות דגימת כדור שלג וערכו ראיונות בשיחות וידיאו. במהלך הריאיון נשאלו הקלינאים על מקרים שבהם נתקלו במחסום שפה עם מטופל וכיצד פעלו במצב זה; מתי, אם בכלל, הם השתמשו בכלי תרגום מכונה ומדוע; וכיצד הם העריכו את איכות התרגום של מתורגמן רפואי לעומת כלי תרגום מכונה. נשאלו גם שאלות על האתגרים של מתן טיפול לאוכלוסיות חלשות, כיצד הטכנולוגיה משולבת כיום בתהליכי העבודה שלהם, וכיצד מחסומי שפה מפריעים לתקשורת מילולית וכתובה כאחד.

ניתוח נתונים: לאחר תמלול הקלטות הראיונות, בוצע קידוד איכותני פרשני ולאחריו החוקרות דנו בנושאים שחזרו על עצמם בראיונות שונים. חלק מהדוגמאות לנושאים אלה הם מחסומי שפה והשפעתם על אופן הטיפול ואיכות הטיפול, מדידת איכות התרגום, אתגרים בעבודה עם אוכלוסיות פגיעות, שילוב כלי MT בעבודה היומיומית של הקלינאים, ועבודה מול מטופלים עם מחסומי שפה באינטראקציות בעלות סיכון נמוך לעומת סיכון גבוה.

**תוצאות**

נמצא כי אתגרי התקשורת איתם התמודדו קלינאים נבעו מחוסר זמן ומשאבים, כמו גם מחסומים תרבותיים ואוריינות בריאות (health literacy). הם גם הביעו דאגה לגבי הדיוק של תרגומים ואחריותיות (accountability) בעת שימוש בתרגום מכונה. כאשר התמודדו עם מחסום שפה, קלינאים הסתמכו על שילוב של מתורגמנים רפואיים, תרגום מכונה והידע שלהם בשפה האחרת. בתגובה לאתגרי תקשורת בין שפות, רופאים פיתחו אסטרטגיות להערכת תקשורת עם מטופלים הכוללות תרגום חוזר, תקשורת לא מילולית ובדיקת הבנת המטופל.

אתגרי תקשורת בין שפות

*זמן ומשאבים מוגבלים.* מגבלות זמן ומשאבים מפריעים לתקשורת בין מטופלים לרופאים ומשפיעים לרעה על איכות הטיפול, אתגרים אשר מחריפים במסגרות רפואיות דלות משאבים. הפתרון האידאלי לתמיכה במצבים אלה הוא שימוש במתורגמנים רפואיים, אך מפאת מחסור בזמן קלינאים לעתים קרובות נמנעים מלהיעזר בשירותי תרגום אלה. הרופאים תיארו כיצד עצם ההתקשרות לשירותי מתורגמנות אורך זמן רב, וכיצד התרגום מאריך באופן ניכר את משך הביקור. בגלל הזמן המוגבל שיש לרופאים עם מטופליהם, עליהם לבחור בקפידה את השאלות למטופל וכך נפגמת איכות הטיפול. רופאים תיארו שלעתים קרובות הם לא יכולים להתקשר למתורגמן עקב משאבים מוגבלים, השעה ביום או בגלל שהמטופל דובר ניב ספציפי או שפה פחות נפוצה אשר אין מתורגמנים זמינים הדוברים אותם. במקרים כאלה, הרופאים צריכים לקבוע מחדש פגישות עם אותם מטופלים או להסתפק באמצעים פחות טובים לתקשורת עימם.

*מחסומים תרבותיים ואוריינות בריאות.* החוקרות מצאו שתקשורת מטפל–מטופל הושפעה לרעה גם במקרים שבהם למטופלים היו שיעורי אוריינות בריאות נמוכים, במיוחד בקרב אוכלוסיות שאינן דוברות אנגלית, ובמקרים שבהם היו מחסומים תרבותיים בין מטופלים לרופאים. בקרב אוכלוסיות דוברות אנגלית, שיעורי אוריינות בריאות נמוכים מובילים בדרך כלל לכך שחולים טועים בלקיחת תרופות, או חוזרים למסגרות רפואיות עם שאלות המשך. באוכלוסיות שאינן דוברות אנגלית, אוריינות בריאות נמוכה מציבה אתגרים נוספים שכן רופאים התקשו לפענח אם השגיאה בתקשורת היא בגלל תרגום שגוי או עקב חוסר היכרות עם ז'רגון רפואי בשפת האם של המטופל.

כאשר לשיעורי אוריינות הבריאות הנמוכה התווספו פערים תרבותיים, איכות הטיפול ירדה באופן דרסטי – חסמים תרבותיים הפריעו לשיטות שהקלינאים השתמשו בהם כדי לבדוק את הבנת המטופלים. החוקרות מציינות את תרבות ה'כן' הנפוצה כדוגמה לאחד המחסומים שיש לו השפעות שליליות על הטיפול: כשמטופלים נשאלים אם הבינו, הם עונים "כן" על כל שאלה, גם אם אין זה המקרה. תופעה זו מתרחשת בחולים גם ללא קשר לנוכחות של מחסום שפה, אך ההשפעות היו בולטות יותר בקרב מטופלים מאוכלוסיות שאינן דוברות אנגלית, שהייתה להם נטייה לספר לרופאים שהם מבינים את תוכנית הטיפול שלהם, אבל אז היו מעלים שאלות במהלך אינטראקציות עם אחיות.

*דיוק ואחריותיות*. נמצא במחקר שכאשר הקלינאים התמודדו עם מחסומי שפה, הם הסתמכו על מתורגמנים רפואיים מוסמכים כאשר הייתה להם גישה למתורגמן, כאשר היה להם מספיק זמן ועבור שיחות עם סיכון גבוה יותר. הם פנו לתרגום מכונה כאשר לא היה אפשר לפנות למתורגמן או כאשר זה נראה יקר מדי בהתחשב בנסיבות. בשני המקרים, הרופאים דאגו לגבי דיוק התרגום, במיוחד כאשר הם לא ידעו את השפה האחרת. הרופאים דאגו גם לגבי אחריותיות – בשיחות עם סיכון גבוה (מקרים שבהם נדרשה הסכמת המטופל) נוכחות מתורגמן נתפסה כחובה. למרות שלא הייתה דרישה חוקית לעשות זאת, שימוש במתורגמן שיתרגם ויאמת את הסכמת המטופל סיפק רובד מגן של אחריות שמערכת MT לא יכולה לספק. נעשה שימוש בתרגום מכונה רק כאשר זו הייתה האפשרות האחרונה ולא הייתה ברירה אחרת.

בשיחות בעלות סיכון נמוך, רופאים בדרך כלל הסתפקו בשיחה חלקית או ב־Google Translate כדי להתגבר על מחסומי שפה. מצבים כאלה כללו שאלות פשוטות ולקיחת היסטוריה של המטופל, שלהם יש אחריות משפטית פחותה משמעותית ועל כן אינן מצדיקות מתורגמן רפואי. השיקול שחזר בראיונות היה הזמן שייקח להתקשר למתרגם לעומת מידת הדחיפות של המצב רפואי של המטופל. ההחלטה אם להתקשר למתרגם הייתה תלויה במידה רבה ברמת הסיכון של הסיטואציה, ואם לקלינאים הייתה היכרות בסיסית עם השפה שבה דיבר המטופל. רופאים רבים פנו ל-Google Translate אם הם הכירו את השפה בה דיבר המטופל אך לא היו מוסמכים בה מבחינה רפואית, והשתמשו בתוכנה רק כדי לאמת מילים רפואיות ספציפיות.

האסטרטגיות לתקשורת בין שפות

קלינאים השתמשו בארבע טקטיקות עיקריות כדי לגשר על פערים בתקשורת בינם לבין המטופלים שלהם:

*ניסוח מחדש ופישוט של מונחים רפואיים*. הרופאים ניסחו מחדש את המונחים והשתמשו בשפה פשוטה כדי לשפר את דיוק התרגום. הם עשו זאת גם עם מתורגמנים רפואיים, שייתכן כי אין להם ידע בז'רגון רפואי מורכב, וגם עם מערכות תרגום מכונה, מכיוון שהקלינאים היו מודעים לכך שכלים כמו Google Translate פותחו למטרות כלליות ולא פותחו לשימוש במסגרות רפואיות באמצעות אוצר מילים רפואי.

*תרגום חוזר.* חלק מהרופאים השתמשו בשיטות תרגום חוזר שבו ביקשו מהמתורגמן לחזור על מה שהוא תרגם למטופל, כדי לוודא שאכן הבין את כוונתם כראוי.

*מחוות, ציורים ותקשורת לא מילולית.* בהיעדר מתורגמנים רפואיים זמינים, רופאים השתמשו במגוון שיטות לשיפור התקשורת עם מטופלים שאינם דוברי אנגלית. השיטות כללו ציורי עט ונייר, עזרים חזותיים ומחוות.

*רמזים חברתיים ושיטת לימוד.* קלינאים הסתמכו על רמזים חברתיים, כמו שפת הגוף והבעות הפנים של המטופלים, כדי להעריך אם הבינו את הנאמר להם. כמו כן, קלינאים השתמשו בשיטת לימוד חוזר (teach-back), בה ביקשו ממטופלים לחזור על המידע שנאמר להם כדי להעריך את מידת הבנתם.

**דיון ועבודה עתידית**

המשתתפים במחקר המציאו אסטרטגיות להתמודדות עם אתגרים מול מחסומי שפה. רוב המשתתפים הביעו עניין בפוטנציאל לתקשורת עם המטופלים שלהם בתיווך MT כל עוד המערכת תוכננה באופן קפדני וראוי לשימוש בסביבה קלינית. בעוד שמערכות MT כלליות אינן מספקות את הצרכים הללו, ממצאי המחקר מוכיחים פוטנציאל לפיתוח מערכות MT שתוכננו במיוחד לשימוש בטוח ואמין בתחום הבריאות. החוקרות מציגות את השלכות הממצאים על עיצוב MT עבור מסגרות רפואיות, ודנות בעבודה עתידית בנושא זה.

תרגום מדויק של שפה רפואית

כשרופאים פנו לכלי תרגום מכונה מסחריים כלליים כמו Google Translate, הם זיהו שהכלים הללו לא נבנו לשימוש במסגרות רפואיות, ועל כן עלה חשש שהם לא יתרגמו במדויק שפה רפואית. תרגום מדויק חשוב במיוחד במסגרות עתירות סיכון כמו בריאות ורפואה, שבהן טעויות עלולות להיות מסכנות חיים. בעוד שמערכות MT עצביות תמיד מספקות את התרגום הסביר ביותר מבחינה סטטיסטית של ביטוי, בסביבות ושיחות בעלות סיכון גבוה לרוב עדיף לא לתרגם כלל מאשר לתרגם תרגום שגוי. מערכות MT לשימוש בתחום הבריאות חייבות להיות מתוכננות במיוחד כדי לתת עדיפות לתרגום מדויק של שפה רפואית. גורם חשוב נוסף שרופאים זיהו לפיתוח מודל עתידי הוא הבטחת תמיכה בניבים שונים של שפות. לא רק שחסרה תמיכה בניבים שונים בכלי MT קיימים, אלא הם אפילו לא מציינים באיזה ניב של שפה הם כן תומכים, או האם ה־MT מייצר שפה בניב עקבי בכלל.

גישה משלימה היא לשלב תרגום מכונה עצבי עם ביטויים מתורגמים באופן מקצועי. בתחומים רבים של הרפואה ישנם נושאי שיחה נפוצים וסטנדרטיים יחסית (כמו היסטוריה רפואית, תסמינים נוכחיים), אשר יכולים להתאים ליישומים מבוססי ביטויים קבועים. יישומים אלה אמנם מגבילים את מגוון הדברים שאפשר לתקשר, אך מבטיחים תרגום מדויק. לכלים המשלבים תרגום מבוסס ביטויים עם MT עצבי יש הפוטנציאל לספק ערובה גבוהה יותר לדיוק, תוך מתן אפשרות לתקשורת גמישה יותר בעת הצורך. השימוש בכלים התומכים ב־MT עשוי להיות בטוח יותר לשימוש עבור רופאים ומטופלים אם יהיה אפשר לראות אילו חלקים של הטקסט תורגמו באופן מקצועי.

הערכה ואישור קפדניים

רופאים ציינו שוב ושוב את החשיבות של כלי תרגום מכונה שנבדק ואושר על ידי גוף מנהל, כגון ועד מנהל של בית חולים או חברה רפואית. אימות כזה מעיד בדרך כלל שכלי נבדק באמצעות ניסוי בקרה אקראי, ושהממצאים פורסמו בכתב עת רפואי שנערך בביקורת עמיתים. שיתופי פעולה בין שירותי בריאות ולמידת מכונה הינם קריטיים לפיתוח סטנדרטים קפדניים ועקביים להערכת מערכות MT חדשות לשימוש קליני.

נחוצה עבודה עתידית כדי להבין כיצד ניתן למדוד את הסיכונים והיתרונות של MT בסביבה קלינית. ממצאי מחקר זה מצביעים על כך שהדבר שהכי חשוב לרופאים הוא לא איכות התרגום לשמה, אלא כיצד כלי תרגום מסוגל לתמוך בתקשורת יעילה יותר בין מטופל למטפל באופן בטוח, ובכך לסייע למטופל ולשמור על בריאותו. בעוד שכלי MT צריכים להיות מוערכים תחילה עבור רמות מקובלות של דיוק תרגום, יש להעריך אותם גם באמצעות ניסויים עם בקרת אקראיות כדי להבטיח שיש להם השפעה מועילה על בריאות המטופלים ולוודא שאינם גורמים נזק עקב תרגום שגוי.

התקדמות מעבר לתרגום אחד לאחד

הפונקציונליות שעומדת בבסיס מערכות תרגום מכונה קיימות היא עיבוד קלט לתרגום מיטבי יחיד לשפה אחרת. הגישה במחקר זה שואפת לכוון מחדש את המטרות של מערכות MT לשימוש קליני, מאופטימיזציה אובייקטיבית לכאורה של איכות התרגום בין שני טקסטים, לעיצוב איכות כוללת של תקשורת בין מטופלים לרופאים. הראיונות עם הקלינאים העלו תובנות לגבי אתגרי תקשורת והטקטיקות שבהם נקטו כשנתקלו באתגרים, ומהן ניתן להציע רעיונות לתכנון תמיכה טכנולוגית שתהיה מתקדמת יותר מכלי המתרגם טקסט אחד לאחד. הנושאים אליהם יש לשים לב בעת פיתוח מערכת כזו מופיעים להלן:

*מתן תשומת לב ומשקל לשליטה משתנה בשפות*. מערכות MT קיימות אינן מביאות בחשבון את מיומנות השפה של המשתמשים בשפת המקור או היעד – לעתים אנשים משתמשים בתרגום מכונה כאשר יש להם ידע כלשהו בשפה השנייה. למשל, קלינאים בעלי בקיאות מסוימת בשפה אך שאינם מוסמכים להשתמש בשפה זו בפרקטיקה הקלינית, עשויים להשתמש ב־MT כדי לחפש מונחים לא מוכרים או לבדוק שוב את התרגומים שלהם. כמו כן, מטופלים בעלי שליטה מוגבלת באנגלית עשויים להבין אנגלית פשוטה אך להתקשות בהבנת ז'רגון רפואי. התחשבות במגוון רחב יותר של מיומנויות השפה של המשתמש פותחת הזדמנויות עיצוב חדשות עבור מערכות MT ותמיכה בשפה בצורה רחבה יותר. לדוגמה, שילוב תמיכת MT בפורטלים של מטופלים יכול לאפשר למטופלים לתרגם רק חלקים מטקסט שהם לא מבינים, או לקלינאים לצפות בקלות בהצעות או תרגומים בזמן שהם כותבים. בכיוון השני, מטופלים ורופאים בעלי ידע דו־לשוני יכולים להציע משוב שימושי למערכת MT כאשר התרגום אינו נכון או מבלבל.

*תמיכה בבדיקת הבנה*. רופאים הדגישו שתקשורת בין־לשונית אינה עוסקת רק בהעברת מידע למטופל, אלא גם בבדיקת הבנתו והצעת הזדמנות להבהיר לפי הצורך. מערכות MT תומכות יותר ויותר בשיחה דו־כיוונית, למשל, תרגום של טקסט דבור ולא כתוב או החלפת שפות מקור ויעד בקלות, אך תכונות אלו אינן מתוכננות להתאים ללימוד חוזר או שאלות הבהרה. עבודה עתידית יכולה לעצב תכונות ספציפיות התומכות בפרקטיקות הקיימות של הרופאים, או להציג אסטרטגיות חדשות לבדיקת הבנת המטופל תוך כדי פעולה.

*אתגרי אוריינות.* כדי שמערכות MT יקדמו שוויון בתחום הבריאות, יש לשקול כיצד לתמוך בחולים עם רמות שונות של אוריינות בשפתם העיקרית, כמו גם בחולים הדוברים שפות שאין להן צורה כתובה. אתגרי האוריינות כוללים גם את היכולת לקרוא הוראות ומידע כתובים וגם אוריינות בריאות. כיוון אחד לתכנון מערכת שתתן מענה לאתגר זה יהיה מערכת תמיכה מולטי־מודאלית, הכוללת תקשורת חזותית ואודיו.

קביעת תקני בטיחות שקופים ושיטות עבודה מומלצות

למרות שהקלינאים היו מודעים לכמה מהמגבלות של מערכות MT (כמו תרגומים לא אמינים של טרמינולוגיה רפואית), המודעות הזו הייתה למרות שקיפות מוגבלת ממערכות MT זמינות. לדוגמה, ביצועי MT משתנים מאוד בין צמדי שפות, בהתבסס על כמות נתוני ההדרכה הזמינים וההשקעה שהושקעה בפיתוח מודלים לכל שפה, אך מידע זה אינו מועבר למשתמשי הקצה של מערכות מסחריות. מעצבים המפתחים כלי MT עבור שירותי בריאות חייבים לשקול דרכים להעביר בבירור את מגבלות המערכת למשתמשיה. נדרשת עבודה נוספת לפיתוח חומרים משולבים והנחיות קונקרטיות כדי לעזור לרופאים להשתמש ב־MT בבטחה. בהינתן הזמן המוגבל של הרופאים, המערכת תציע באופן אידיאלי התראות כאשר נראה שמשתמש מפר את ההנחיות הללו, עם הצעות אינטראקטיביות כיצד לפתור את הבעיה ולשפר את איכות התרגום.

מגבלות ושיקולים אתיים

למרות שיש פוטנציאל משמעותי לשיפור בכלי MT כדי לתמוך בתקשורת בין לשונית במסגרות קליניות, אין זה ללא סיכונים ומגבלות. נחוצה תמיכה שפה נוספת נחוצה במרפאות דלות משאבים ועבור שפות בעלות משאבים נמוכים, שבהן יש סיכוי נמוך יותר שמתורגמן אנושי יהיה זמין. עם זאת, קיים סיכון שהמאמץ הכרוך בהטמעת טכנולוגיה חדשה ישאב עוד יותר משאבים ותשומת לב ממטופלים בעלי שליטה מוגבלת באנגלית, ורק יחריף את הפערים באיכות הטיפול במקום לצמצם אותם. יתר על כן, כל שגיאה שמערכת MT תעשה תשפיע באופן לא פרופורציונלי על מטופלים שמדברים שפות בעלות משאבים נמוכים יותר. סיכון זה מדגיש את החשיבות של שיטות הערכה קפדניות, איטרטיביות ורב־גוניות. כל מערכת MT לשירותי בריאות חייבת לעבור הערכת מהימנות ודיוק, כמו גם הערכת תועלת לבריאות המטופל. הערכות אלו חייבות להתבצע בזהירות כדי למנוע עומס או סיכון מיותר עבור המטופלים והרופאים המעורבים.

תרגום שפות אינו האתגר היחיד שעמו מתמודדים קלינאים בתקשורת בין־לשונית עם מטופלים. נושא בולט שעלה בראיונות היה התמודדות עם קשיים במתן טיפול מותאם תרבותית ובניית אמון עם מטופלים בעלי הבדלים תרבותיים. קשיים אלה מחריפים לנוכח מחסומי שפה אך אינם בהכרח נפתרים על ידי תמיכת מתורגמן בלבד. על מפתחי MT לשירותי בריאות מוטלת האחריות לתקשר בצורה ברורה את המגבלות של מערכות MT, ולהבטיח שהן לא ייראו כפתרון לגישור על הבדלי תרבות או תחליף להשקעה במשאבים כמו חינוך לגבי טיפול מותאם תרבותית.

לבסוף, בעוד שתובנות הקלינאים שעלו במדגם זה מציעות כיוונים חשובים לעבודה עתידית בפיתוח MT עבור שירותי בריאות, גם נקודות מבט של המטופל חשובות ביותר לעבודה זו. עבודה עתידית חייבת לעסוק באופן משמעותי במטופלים כדי להבין כיצד, אם בכלל, הם רוצים ש־MT יהיה חלק מהטיפול בהם.

**סיכום והמלצות**

תקשורת בין־לשונית בין רופאים למטופלים יכולה להשפיע רבות על איכות הטיפול. מערכות תרגום מכונה הן פתרון בעלות נמוכה שיכול להתמודד עם אתגרים שיש לרופאים בעת אינטראקציה עם מטופלים בנוכחות מחסומי שפה. מערכות כאלה, כאשר הן אמינות ובעלות שקיפות לגבי השימוש בהן, יכולות לשפר את איכות הטיפול בחולים במסגרות חסרות משאבים, שבדרך כלל הגישה שלהן למתורגמנים רפואיים מוגבלת. אפילו במסגרות רפואיות עתירות משאבים, מערכות תרגום מכונה יכולות להשלים את שירותי התרגום הקיימים על ידי עבודה עם מתורגמנים רפואיים שלחלקם חסר ידע בז'רגון רפואי מורכב.

המלצות המחקר הן שמערכות תרגום מכונה ישופרו בהיבטים הבאים: לתת את הדעת על הבדלי ניבים; לשלב תרגום מכונה עצבי עם ביטויים מתורגמים באופן מקצועי; להתקדם מתרגומים אחד לאחד אל עבר תקשורת איכותית כוללת; לעודד משוב מערכת לבדיקת הבנה; לקחת בחשבון את רמות האוריינות של מטופלים; להגדיר תקני בטיחות שקופים והצעות אינטראקטיביות לרופאים ולמטופלים. שיפור מערכות MT באופנים האלה יוכל לשרת טוב יותר אוכלוסיות מטופלים מגוונות ולצמצם פערים באיכות הטיפול.

**מאמר 4: פיתוח תרגום מכונה סווהילית קונגולזית לצורך סיוע הומניטרי במגפת הקורונה**

במאמר זה מתוארים המאמצים של מתרגמים מארגון Translators Without Borders (TWB) ליצור מערכת של תרגום מכונה עצבי בין השפות סווהילית קונגולזית לצרפתית, במטרה לשפר את שטף העבודה של תרגום הומניטרי בזמן מגפת הקורונה.

סווהילית היא שפת מאקרו המדוברת במזרח אפריקה ובערך 100 עד 150 מיליון איש דוברים את השפה. זוהי השפה הרשמית בטנזניה ובקניה, שם היא מכונה סווהילית חוף (Coastal Swahili), והיא אחת מארבע שפות לאומיות ברפובליקה הדמוקרטית של קונגו (Democratic Republic of Congo - DRC), שם היא מכונה סווהילית קונגולזית (Congolese Swahili). סווהילית חוף וסווהילית קונגולזית נבדלות באופן משמעותי מבחינת אוצר המילים, הדקדוק והמבנה. תרגום מכונה המבוסס אך ורק על ניב החוף עלול להיות לא יעיל במסירה או קבלה של מידע רגיש, ועל כן עלה צורך בפיתוח תרגום מכונה ספציפי לדיאלקט כדי לסייע למאמצי סיוע הומניטריים ב־DRC.

ארגון TWB מתמחה בסיוע לאנשים שנפגעו ממשבר לקבל מידע חיוני בשפתם. תקשורת יעילה היא בעלת חשיבות מרבית כדי להבטיח שאנשים יבינו סיכונים בריאותיים ויידעו כיצד לשמור על הבטיחות האישית שלהם ושל משפחותיהם. TWB מתרגם מסרים חיוניים לשפות מקומיות ועובד עם גורמים נוספים לפיתוח כלים כדי לספק לקהילות גישה למידע ולשירותים העונים על צרכיהם. באוקטובר 2020, TWB השיק צ'אטבוט רב־לשוני העונה לשאלות בנושא נגיף הקורונה הממנף הבנת שפה טבעית כדי לאפשר למשתמשים לשאול שאלות במילים שלהם ולקבל תשובות רלוונטיות בשפתם.

טכנולוגיות שפה, כמו תרגום מכונה, ממלאות תפקיד חשוב בתגובה למשברים ובהגדלת היכולת להעביר מידע קריטי ומסרים מרכזיים בשפות שונות במהירות ובקנה מידה גדול. אנשים שנפגעו במשבר יכולים לגשת לתוכן בשפות מקומיות ממקור ראשון דרך ערוצים שונים כגון אתרי אינטרנט, מקורות חדשותיים ומדיה חברתית. בכיוון ההפוך, ניתן לעשות שימוש בשאלות המשתמשים ובתגובות שלהם כדי לסייע לתוכניות סיוע לנתח מגמות המתחוללות בשטח ולענות טוב יותר על הצרכים של האנשים המושפעים. עבור שפות מיעוטים שבהן משאבי התרגום מוגבלים, MT יכול גם לעזור לתקנן מונחים ולשפר את איכות התרגום.

המשימה של יצירת MT המתמחה בניב קונגולזי מאתגרת במיוחד בשל המשאבים הנמוכים המוקדשים לשפה. באינדקס של Joshi et al. (2020) המדרג שפות במונחים של משאבים, הניב מאופיין בתור "Left Behind". על פי הגדרתם, קשה מהותית להקים בסיס דיגיטלי לשפה עם משאבי הנתונים הקיימים. אם לא מטפלים בצוואר הבקבוק של מחסור בנתונים, אי אפשר לפתח כלים שימושיים כמו MT שיכולים לשרת את המתרגמים של שפה.

לחוקרים היו ארבע מטרות בעבודתם: לאצור מקורות מחקר ומידע ציבורי לשימוש בפיתוח MT עבור סווהילית קונגולזית; לתאר את מערכי הנתונים שהם יצרו והרכיבו לעבודה זו; לחקור את השימוש במתודולוגיות בעלות משאבים נמוכים כדי לספק מודלים של תרגום מכונה דו־כיווני אופטימלי; להעריך את שמישות ה־MT לתגובה הומניטרית.

**תיאור הנתונים**

נתונים מקבילים הם המרכיב העיקרי לאימון מערכות MT. נתונים מקבילים מכונים גם ביטקסט (bitext), קבוצה של משפטים עם התרגומים שלהם. כדי לבנות מודל דו־כיווני של סווהילית קונגולזית וצרפתית, החוקרים היו צריכים גישה לנתונים מתורגמים בשפות אלו. הם אספו נתונים מקבילים בצמד שפות זה משלושה מקורות:

1. **ערכות Gamayun.** ערכות אלה הינן נקודת מוצא לפיתוח קורפוסים של אודיו וטקסט עבור שפות עם מעט נתוני שפה קיימים מראש, או ללא נתונים כלל. משפטי מקור עבור קורפוס זה נבחרים תוך בדיקה שהם מייצגים שפה יומיומית ואינם שייכים לתחום ספציפי. החלק שפורסם עם מאמר זה הוא קבוצה של 25,302 משפטים בצרפתית שתורגמו לסווהילית קונגולזית.
2. **TICO-19 translation memories.** תרגומים אלה נאספים כדי לסייע למתרגמים וכדי לבנות מדדי MT ב־36 שפות. כל סט הורכב מ־3,071 משפטים מתורגמים בנושא הקורונה שאיכותם נבדקה.
3. **תרגומים פנימיים של TWB**. רוב תוכן המקור שנאסף הוא מבקשות תרגום של ארגונים לא ממשלתיים המסייעים בסעד הומניטרי ב־DRC.

המשאב החיצוני הפתוח היחיד שנמצא במיוחד עבור סווהילית קונגולזית הוא קורפוס JW300. בנוסף, החוקרים אספו נתונים בסווהילית שאינה דיאלקטית כדי שיוכלו לבצע לימוד העברה, ואספו משפטים המשלבים צרפתית ואנגלית.

**יצירת נתונים סינתטיים**

החוקרים כללו בניסויים שלהם גם נתונים בסווהילית חוף המשולבת עם אנגלית. כדי שיוכלו להשתמש בהם לאמן את הדגמים שלהם בשילוב עם צרפתית, הם השתמשו במודל תרגום מכונה כדי לתרגם את המשפטים באנגלית לצרפתית.

טכניקה נוספת לשיפור MT היא אימון בפיקוח למחצה, תהליך הכולל הרחבת נתוני האימון עם נתונים חד־לשוניים בשילוב עם התרגום החוזר שלהם. עבור הניסויים במחקר זה, החוקרים תרגמו אוטומטית 766,398 משפטים סוואהיליים חד־לשוניים שמקורם באתרי ויקיפדיה וחדשות לצרפתית באמצעות מודל שהוכשר במהלך העבודה הזו תוך שימוש בשאר הנתונים הזמינים.

**נתוני בדיקה**

החוקרים השתמשו בחמש ערכות בדיקה לביצוע הערכות אוטומטיות ואנושיות של המודלים: 1,000 משפטים מערכות Gamayun, 500 משפטים מסט TICO19, 2,478 משפטים ממערך JW300, 100 הודעות שנשלחו על ידי משתמשים משיחות צ'אטבוט ועשרה טקסטים של תגובות מוכנות מראש של הצ'אטבוט. שתי קבוצות המשפטים הראשונות נדגמו באקראי ממערכי הנתונים אליהם הם שייכים.

**פיתוח מודלים**

החוקרים אימנו את מודל תרגום המכונה העצבי על בסיס נתונים של טקסט סווהילי וצרפתי בקונגו, שנאסף ממקורות שונים. כדי לשפר את ביצועי המודל השתמשו החוקרים בהגדלת נתונים (data augmentation). טכניקה זו כרוכה ביצירת דוגמאות אימון חדשות על ידי שינוי הדוגמאות הקיימות בדרכים שונות, כגון שינוי סדר המילים או החלפת מילים במילים נרדפות. בכך הם הצליחו להגדיל את כמות נתוני האימון הזמינים למודל, מה ששיפר את יכולתו לתרגם סווהילית קונגולזית לצרפתית. החוקרים השתמשו גם בטכניקת attention mechanism, המסייעת למודל להתמקד בחלקים הרלוונטיים ביותר של טקסט הקלט בעת יצירת תרגום הפלט. טכניקה זו שיפרה את הדיוק של התרגומים ואף הפכה אותם לרהוטים יותר. הטכניקה השלישית בה השתמשו החוקרים נקראת beam search, שיטה להפקת תרגומים מרובים ולבחירת הטוב ביותר על סמך פונקציית ניקוד. השיטה שיפרה את איכות התרגומים והפכה אותם טבעיים יותר.

**תוצאות**

הערכה אוטומטית

החוקרים השתמשו בשיטות שונות כדי לשפר את מערכת תרגום המכונה עבור סווהילית קונגולזית לצרפתית. השיטה הראשונה שבה השתמשו הייתה למידת העברת דיאלקטים. הם אימנו את מערכת תרגום המכונה על סמך נתונים מניבים אחרים של סווהילית, בנוסף לנתוני הסווהילית קונגולזית שהיו להם. כך המערכת למדה יותר על המבנה ואוצר המילים של סווהילית באופן כללי, מה ששיפר את יכולתה לתרגם סווהילית קונגולזית לצרפתית.

השיטה השנייה בה השתמשו הייתה למידה בפיקוח למחצה. החוקרים אימנו את מערכת תרגום המכונה באמצעות שילוב של תרגומים מתורגמים על ידי אדם ושל תרגומים שתורגמו על ידי מכונה. על ידי שימוש בשני הסוגים המערכת למדה מכמות גדולה יותר של נתונים, מה ששיפר את יכולתה לתרגם בין השפות.

שתי השיטות התגלו כיעילות בשיפור איכות מערכת התרגום המכונה. הן הצליחו לשפר את ביצועי המערכת ב־1.5 עד 2.4 נקודות BLEU (BiLingual Evaluation Understudy – מבחן הערכה אוטומטי לתרגום מכונה) מסווהילית קונגולזית לצרפתית, וב־1.4 עד 3.5 נקודות BLEU מצרפתית לסווהילית קונגולזית.

הערכה אנושית

הערכת האיכות של MT חיונית כדי להעריך את השימושיות שלה בעולם האמיתי. החוקרים תכננו את הגדרות ההערכה הידניות שלהם כדי לקבל תובנות לגבי הביצועים של דגמי ה־MT שפיתחו בהקשר של צ'אטבוט העונה על שאלות בנושא נגיף הקורונה. הצ'אטבוט של TWB הפרוס ב־DRC נקרא Uji – עוזר וירטואלי מבוסס בינה מלאכותית המשיב לשאלות, מתעד חששות ונותן מענה למשתמשים בצרפתית, סווהילית ולינגלה.

בפרויקט היו שתי משימות תרגום עיקריות. הראשונה הייתה תרגום של תשובות שנאספו באופן מקצועי לשאלות נפוצות הקשורות לקורונה. תוכן זה נכתב במקור בצרפתית והיה צריך להיות מתורגם לסווהילית. המשימה השנייה בפרויקט הייתה ניתוח של שאלות לא מסווגות, תלונות ומידע אחר הניתן לניטור ממשתמשי הצ'אטבוט דוברי סווהילית.

במקרה השימוש הראשון, החוקרים ביצעו הדמיה של עריכת פלט MT על ידי בלשן כדי ליצור תרגומים נכונים (post editing). הם בחנו את טיב הפלט על ידי בדיקה של רמת המאמץ הנדרשת כדי לתקן את טעויות המכונה, ובחינה של סוג הטעויות שיש לתקן ולערוך. במקרה השימוש השני, הם העריכו ישירות את הפלט של תרגום המכונה, ללא עריכה.

*הערכת post editing:* החוקרים בחרו עשר תגובות של הצ'אטבוט שנכתבו במקור בצרפתית ותורגמו לסווהילית קונגולזית. התשובות היו בנושאים שונים כמו חיסונים, מיתוסים ווריאנטים חדשים של הנגיף. התגובות כללו בממוצע 2.3 משפטים ו-45 מילים. התוצאות הראו הסכמה גבוהה מהממוצע בין המעריכים לגבי סוגי הטעויות שנעשו על ידי המכונה. השגיאות הנפוצות ביותר היו מילים חסרות, הופעה של מילים צרפתיות בתרגום לסווהילית קונגולזית שגיאות בצורות רבים ויחיד.

*הערכה ישירה:* החוקרים השתמשו בקבוצה של מתרגמים על מנת לקבל הערכה אנושית שתשפוט עד כמה תרגומי המכונה (ללא עריכה) מעבירים את המסר העיקרי של טקסט המקור ועד כמה התרגומים היו טובים בסך הכול. למתרגמים הוצג טקסט המקור יחד עם טקסט היעד והם נשאלו שלוש שאלות פשוטות: האם תרגום זה מעביר את המשמעות העיקרית של טקסט המקור? (כן, בערך, לא); עד כמה התרגום טוב? (0 = גרוע מאוד; 10 = מושלם); במקרה של ציון נמוך (0-4) בשאלה הקודמת, אנא הסבירו מדוע תרגום זה אינו טוב. (סוף פתוח).

החוקרים בחרו להתמקד בהבנת המסר המרכזי (שאלה 1) ובאיכות הכללית (שאלה 2) כדי להעריך את התאמת מערכת תרגום המכונה למטרה שלשמה נוצרה, ולהשמיט את מדדי הדיוק והשטף הנפוצים בהערכת תרגום. החוקרים מתארים כי הם הרגישו שמדדים אלה אקדמיים מדי למטרותיהם: השימוש שייעדו לתרגום מכונה שפיתחו התמקד בהשגת הבנה כללית של תוכן שנשלח על ידי משתמשים, ועל כן פחות חשוב שלתרגומי היעד יהיה דקדוק מושלם או רהיטות גבוהה.

כל הטקסטים שתורגמו, נכתבו על ידי משתמשים ששלחו שאלות הקשורות לקורונה. מכיוון שטקסטים אלה היו שאלות שהוגשו בעיקר דרך WhatsApp, הטקסטים היו קצרים, בממוצע של 11 מילים ו־63 תווים.

התוצאות הראו כי בממוצע, הטקסטים דורגו בציון 6.3 מתוך 10, עם סטיית תקן של 3.0. חשוב לציין שניתנה הגדרה מילולית רק לחלק התחתון ולחלק העליון של הסולם, ולכן נדרשה פרשנות אישית מסוימת לשלבים שביניהם, פרמטר המשתנה ממתרגם למתרגם. באופן כללי לא היה קשר בין איכות ואורך טקסט, אבל זה יכול להיות מוסבר בחלקו על ידי ההבדל הקטן בין הטקסטים הקצרים ביותר (37 תווים) והארוכים ביותר (144 תווים). היה קשר חזק יותר בין איכות התרגום לבין האם המסר העיקרי מועבר. נתון זה אינו מפתיע, מכיוון שבאופן טבעי המסר העיקרי מועבר טוב יותר כאשר התרגומים טובים יותר. המעריכים דיווחו שלפחות חלק מהמסר המרכזי של טקסט המקור הועבר ב־75% מהמקרים, ושב־48% מהתרגומים המסר המרכזי הועבר לגמרי.

**מסקנות והמלצות**

מאמר זה מתאר מקרה בוחן של פיתוח תרגום מכונה של שפה דלת משאבים, סווהילית קונגולזית. החוקרים סיפקו תוצאות ניסוי מפורטות על השימוש בלמידה של העברת דיאלקט ובמתודולוגיות למידה מפוקחת למחצה. הערכה אוטומטית בין מערכי בדיקה שונים הראתה שיפור של 1.5 עד 2.4 נקודות BLEU מסווהילית קונגולזית לצרפתית ו-1.4 עד 3.5 נקודות BLEU מצרפתית לסווהילית קונגולזית באמצעות שימוש בשיטות כאלה.

תוצאות מבחני ההערכה על ידי מתרגמים אנושיים מדגימות פוטנציאל לשיפור זרימות העבודה של תרגום בתרחישי סיוע הומניטרי. החוקרים פרסמו אפליקציית דמו למתרגמים הבנויה על המודלים שהשתמשו בהם כדי לשרת הן את האוכלוסייה הכללית הן את המתרגמים. הם פרסמו גם את המשאבים שנבנו ונאספו לצורך המחקר, כדי שחוקרים אחרים יוכלו להמשיך ולקדם מחקר NLP ופיתוח טכנולוגיית שפה עבור סווהילית קונגולזית.

החוקרים כותבים כי הם משלבים בעבודתם את המודלים שפיתחו וכי המטרה העיקרית שלהם היא לתת כלי זה לקהילת המתרגמים של TWB ולהמשיך ולפתח את המודלים עם המשוב שיתקבל מבלשנים ומתרגמים לא מקצועיים המגיעים מרקע ומניסיון שונה.

**דיון**

בעבודה זו הוצגו ארבעה מאמרים העוסקים בשימוש בתרגום מכונה במסגרות רפואיות, ארבעתם מציגים מחקרים אמפיריים שונים בתחום. המאמר הראשון (Taira et al., 2021) הוא מחקר כמותי העוסק בהערכת הדיוק של Google Translate לתרגום מכתבי שחרור מחדר מיון עבור חולים בעלי שליטה מוגבלת באנגלית. המאמר השני (Lee et al., 2023) גם הוא מחקר כמותי, והוא בודק את איכות הביצועים של שלושה יישומי תרגום מכונה מסחריים לצורך תרגום בעל פה של שיחות בין אנשי צוות רפואי ומטופלים. המאמר השלישי (Mehandru et al., 2022) הוא מחקר איכותני העוסק בפתרונות שבהם נוקטים קלינאים כאשר הם מטפלים בחולים בעלי שליטה מוגבלת באנגלית, ומסיק מהן מסקנות לגבי דרכי שימוש בטוחות ואמינות יותר בתרגום מכונה בסביבה רפואית. במאמר הרביעי (Oktem et al., 2021) בוצע מחקר כמותי בתחום הבלשנות החישובית והוא מתאר את מאמציהם של מתרגמים מארגון Translators Without Borders לפתח כלי תרגום מכונה בין השפות סווהילית קונגולזית וצרפתית לצורך מענה הומניטרי בזמן מגפת הקורונה. למרות, ואולי אף בזכות ההבדלים בין המאמרים, ניתן לדון בסוגיות שונות העולות מהם.

**סוגיה ראשונה: שימוש בכלי תרגום מכונה כפי שהם לעומת שימוש חלקי או מושכל**

בשני המאמרים הראשונים ניכרת גישה יחסית מחמירה בנוגע לשימוש בתרגום מכונה כאשר איכות התרגומים, לדעת החוקרים, אינה טובה דיה. במאמר הראשון של Taira et al. (2021), למרות ציונים יחסית טובים שהתקבלו בהערכת איכות התרגומים וברמת פגיעה אפשרית במטופלים בעקבות תרגום שגוי, המלצת החוקרים בסופו של דבר הייתה לא להשתמש כלל בתרגום מכונה. ברצוני להדגיש נתון מהמאמר: "שיעורי הדיוק שנמצאו כפי שהוערכו על ידי מתנדבים מהקהילה עמדו על 6% אי־דיוקים בספרדית ו־18% אי־דיוקים בסינית, *אחוזים כמעט זהים לרמות הדיוק של מתרגמים מקצועיים*". למרות נתון זה, המסקנה היא שהשימוש ב־Google Translate טרם בטוח ואמין במחלקה לרפואה דחופה. כמו כן, רמת הדיוק שהתקבלה במאמר זה עבור תרגום מאנגלית לספרדית עמדה על 94%, ודיוק התרגום לטאגלוג עמד על 90%. אך החוקרים קובעים כי מכיוון שהדיוק של GT משתנה באופן מהותי לפי שפה, הוא עדיין אינו כלי אמין לשימוש בקליניקה. כלומר למרות שישנן שפות שבהן התרגום טוב דיו ומתקרב באיכותו לתרגום של מתורגמנים מקצועיים, החוקרים ממליצים לא להשתמש בו באופן גורף, בגלל שבשפות אחרות רמות הדיוק שלו נמוכות יותר. נראה כי עולה גישה של "הכול או כלום" – מכיוון שהכלי אינו מתרגם היטב בחלק מהשפות, ההמלצה היא שלא להשתמש בו בכלל.

גם במאמר השני (Lee et al., 2023) המלצת החוקרים היא שלא להשתמש בתרגום מכונה עבור תרגום בעל פה של שיחות בעלות סיכון נמוך, לאחר שאף אחד מהיישומים שנבדקו במחקר לא עבר את סף האי־נחיתות. זאת למרות שהם ציינו במגבלות המחקר כי ייתכן שחלק מהבעיות שבגללן היישומים קיבלו ציונים נמוכים במחקר, יכולות להיפתר בקלות במציאות. למשל, אדם המשתמש באחד מהיישומים ורואה שיש בעיה בתמלול דבריו יכול להתחיל את התרגום מחדש ולחזור על מה שאמר לאט וברור, יכול לתקן באופן ידני את התמלול כדי שטקסט המקור יהיה ללא שגיאות, ויכול גם להקליד את הטקסט מלכתחילה ולהשמיע את פלט האודיו למטופל, וכך לעזור למכונה לייצר תרגום שיהיה איכותי ומובן יותר. שוב עולה גישת "הכול או כלום", בצורת המלצה גורפת לא להשתמש ביישומי תרגום מכונה כלל למרות שבאמצעות התערבות קלה מצד המשתמש, אפשר להפיק תרגומים מוצלחים יותר מאלה שהתקבלו במחקר.

במאמר השלישי (Mehandru et al., 2022) ניכרת מגמה אחרת. כותבות המאמר מתייחסות לחסרונות של תרגום מכונה, כמו בעיות דיוק, קושי להעריך את התרגום כאשר המשתמש אינו דובר את שפת היעד, ובעיית האחריותיות בעת שימוש בתרגום מכונה. אולם הן לא מציעות שלא להשתמש כלל בתרגום מכונה בגלל בעיות אלה, אלא מציגות דרכים בטוחות יותר להשתמש בתרגום מכונה שיוכלו להוביל לתרגומים מדויקים יותר ובד בבד מציעות דרכים לשיפור תרגום מכונה כך שיהיה אפשר להסתמך על כלי זה יותר בעתיד. ההצעות שלהן לפיתוח תרגום מכונה מסתמכות על הראיונות עם הקלינאים והן מעשיות וקונקרטיות, דבר שאמור לא רק לשכלל את ביצועי המערכת ולדייק אותם לשיחות רפואיות בעלות סיכון נמוך, אלא גם להגביר את האימוץ שלה בקרב צוותים רפואיים. הצעות אלה כוללות פיתוח מודל ספציפי לשימוש בתחום הבריאות שיתמוך בתרגום שפה רפואית וישלב יחד טכנולוגיות קיימות (תרגום ביטויים קבועים בשילוב NMT) כדי להגיע למוצר הכי אמין ובטוח שאפשר כרגע בתחום זה; הערכה ואישור קפדניים על ידי גוף רשמי מתחום הבריאות; פיתוח כלי שיתמוך בתקשורת במובן רחב יותר מאשר תרגום טקסט אחד לאחד, שיכלול התייחסות לשליטה משתנה בשפות, בדיקת הבנה ואתגרי אוריינות; הצגת יכולות המערכת בשקיפות מלאה ושימוש בה בדרכים דינמיות שיאפשרו תרגום מדויק יותר.

חשוב להזכיר כי הצעות אלו מכוונות לשימוש בתרגום מכונה במסגרת ספציפית מאוד – שיחות רפואיות בעלות סיכון נמוך. החוקרות מתייחסות למגבלה שכנראה תמיד תהיה לתרגום מכונה, והיא שלעולם לא יוכלו להשתמש בכלי זה לשיחות בסיכון גבוה. במאמר נכתב כי "בשיחות בעלות סיכון גבוה עדיף לא לתרגם כלל מאשר לתרגם תרגום שגוי" – בראש ובראשונה בגלל הסיכון להזיק למטופלים, וסיבה משנית היא בגלל שאם חלילה מישהו יינזק, מישהו אחר יצטרך לשאת באחריות לכך ואי אפשר להטיל אחריות על מכונה. כותבות המאמר מודעות לחסרונות MT ואינן מתעלמות מהם, אך במקום להצביע על החסרונות כסיבה לא להשתמש בתרגום מכונה עד שתשתפר האיכות, הן מציעות רעיונות לשיפור בעצמן ומראות שלא חייבים להשתמש בכלי כפי שהוא ובכל מצב, אלא אפשר להיעזר ב־MT בתנאים מסוימים ובשימוש מושכל. תרגום מכונה אינו חייב להיות מושלם בשביל לסייע ולתמוך בתקשורת בין מטפלים למטופלים, כי הוא לא הגורם היחיד שמתווך ביניהם אלא רק כלי עזר.

אחד החידושים במאמר זה הוא הצגת כל האפשרויות שבהן נוקטים קלינאים כאשר הם נתקלים במחסומי שפה, כמו שימוש בעט ונייר ובמחוות ידיים. החוקרות אינן משוות תרגום מכונה למתורגמנות רפואיות ממבט שבו שתי אפשרויות אלו הן היחידות שקיימות, ובהיעדר תרגום מכונה בהכרח תהיה מתורגמנות רפואית מקצועית; אלא המתורגמנות הרפואית מקבלת את אותו המשקל שמקבלים הפתרונות האחרים שמתרחשים במציאות. מתוך הסתכלות ריאלית זו על התמונה הרחבה, אפשר להבין מדוע הן לא דוחות תרגום מכונה שאינו טוב דיו לשימוש בפני עצמו אלא מציעות דרכים בטוחות להשתמש בו ביומיום עבור סיטואציות ספציפיות, מתוך ידיעה שה־gold standard של מתורגמנות מקצועית לא תמיד זמין או ישים.

גם במאמר הרביעי (Oktem et al., 2021) ניכרת גישה של שימוש מושכל בתרגום מכונה. עורכי המחקר מסתכלים על הנתונים מזווית אחרת לגמרי: הם מתמקדים בהיבט של הבנת המסר הכללי, בשלמותו או בחלקו, ועל אף שהציונים אינם מאוד גבוהים עבור הבנת המסר המתורגם בשלמותו, הם משתמשים בכלי שפיתחו מכיוון שהוא יכול לסייע להם בשטף העבודה. הם מכירים בכך שהאלטרנטיבה השנייה היא תרגום טקסטים בעצמם ללא כל סיוע מגורם נוסף, ולכן מכירים בהקלה שתרגום מכונה יכול לספק להם, גם אם הוא חלקי בלבד. הסיוע בשטף עבודת התרגום עוזר להם לתרגם יותר חומרים וכך לסייע ליותר תושבים בקונגו.

ייתכן שניתן לייחס את ההבדלים בגישות לדיסציפלינות מהן מגיעים החוקרים. שני המאמרים הראשונים (Taira et al., 2021; Lee et al., 2023) נכתבו על ידי צוותים רפואיים, המאמר השלישי (Mehandru et al., 2022) נכתב על ידי חוקרות מתחום מדעי המחשב, והמאמר הרביעי (Oktem et al., 2021) נכתב על ידי מתרגמים. הצוותים הרפואיים אינם מתמחים בתרגום ובשפות וייתכן שמסיבה זו הם אינם מציעים דרכים להשתמש באופן חלקי ומושכל בתרגום מכונה. כמו כן, הצוותים הרפואיים ככל הנראה מודאגים יותר מהפן של האחריותיות ולכן אינם מתפשרים על כלי שהוא פחות ממושלם. כותבות המאמר השלישי מציעות דרכים לשיפור תרגום מכונה כי הן מגיעות מעולם של פיתוח, הן יודעות שטכנולוגיות משתכללות מיום ליום ושמות את הדגש לא על "מתי" ו"האם" תרגום המכונה ישתפר, אלא מדגישות את ה"איך". המתרגמים, כותבי המאמר הרביעי, רגילים להשתמש בתרגום מכונה בשטף העבודה שלהם ומכירים את הדרכים לנצל את היתרונות של הכלי למרות חסרונותיו. כמו כן, הם פיתחו את הכלי במצב חירום הומניטרי שבו עומס ודחיפות העבודה הצריכו תגבור לסיוע בעבודת התרגום ובמצב כזה עדיף כלי שיסייע באופן חלקי על פני האלטרנטיבה, תרגום ללא סיוע כלל.

**סוגיה שנייה: הערכת איכות לאור הסקופוס**

הסוגיה השנייה שעולה מעיון בארבעת המאמרים יחדיו היא – מהי מטרת התרגום מכונה? תיאוריית הסקופוס של ורמיר (Vermeer, 2000) היא תיאוריה פונקציונליסטית הגורסת כי לכל תרגום יש "סקופוס" – מונח טכני המגדיר את מטרת התרגום או ייעודו. על פי התיאוריה, המתרגם והלקוח צריכים לקבוע מראש מהו הסקופוס, כלומר מהי מטרת התרגום של הטקסט המדובר ואיזו השפעה רצוי שתהיה לו ביעד שלשפתו מתרגמים אותו. צורות שונות של תרגום מקובלות כל עוד המתרגם מודע למעשיו ולהשלכות בחירותיו – מה תהיה השפעת הטקסט בתרבות היעד או אצל לקוח הקצה, עד כמה השפעת התרגום תהיה שונה מהמקור או דומה לו. במאמרים שהוצגו, ניתן לבחון את מטרת השימוש בתרגום מכונה לפי שני פרמטרים: אופן השימוש בכלי תרגום מכונה (עם התערבות או בלעדיה), ואיכות התרגום של הכלים (תרגום מדויק לעומת תרגום טוב דיו).

אופן השימוש בתרגום מכונה

במאמר הראשון בדקו היתכנות של הסתמכות בלעדית על תרגום מכונה, כזו שאפשר להשתמש בה ללא כל התערבות מצד המשתמש וללא שימוש בתרגום מקצועי. גם במאמר השני בדקו אם זה אפשרי להשתמש בתרגום מכונה לתרגום בעל פה ללא שימוש באסטרטגיות תקשורת אחרות וללא ביצוע שינויים והתאמות בטקסט מצד המשתמש. אם המטרה, או לפחות חלק ממנה, של השימוש בתרגום מכונה היא שאפשר יהיה להשתמש בה ללא כל התערבות או מאמצים מצד המשתמשים, ושהיא תחליף ולא רק תשלים תרגום מקצועי, אזי קשה שלא להסכים עם מסקנות החוקרים במאמרים הראשונים: "Google Translate עדיין אינו כלי אמין לשימוש בקליניקה" (Taira et al.2021), "עד שהאיכות של מתורגמנות MT תשתפר משמעותית, על הרופאים להבטיח טיפול בטוח, יעיל והוגן על ידי עבודה עם מתורגמנים רפואיים מקצועיים במידת האפשר" (Lee et al., 2023). איכות התרגום של רוב תרגומי המכונה כיום היא כזו שלרוב מחייבת שימוש מודרך ומושכל והתערבות מצד המשתמש על מנת להפיק תרגומים מדויקים.

במאמר השלישי ציינו כי מטרת התרגום מכונה היא לתמוך ולסייע לתקשורת בין מטפלים למטופלים, אך לא כחלופה בלעדית ויחידה לאסטרטגיות תקשורת אחרות אלא ככלי עזר. הדבר עולה גם מתיאורי הרופאים את דרכי ההתמודדות שלהם מול מחסומי שפה, שכללו הסתמכות על שילוב של מתורגמנים רפואיים, תרגום מכונה וידע קודם בשפות אחרות, וגם מהמלצות החוקרות לעבודה עתידית, המפרטות דרכים שונות לשיפור ושכלול כלי תרגום מכונה כך שיהיו בטוחים יותר, ולא המלצות על הימנעות מכלים אלו. במאמר הרביעי מטרת התרגום מכונה היא לסייע בשטף העבודה, וגם כאן אין המטרה להחליף לגמרי את המתרגמים המקצועיים בעבודתם ההומניטרית אלא רק לסייע להם. הם ציינו במאמר כי הם משלבים בעבודתם את המודלים שפיתחו לתרגום מכונה של סווהילית קונגולזית וצרפתית, וכי המודלים יכולים לשפר את זרימת העבודה של תרגום בתרחישי סיוע הומניטרי. כלומר בשני המאמרים האחרונים, מסתכלים על תרגום מכונה כמשלים, אך לא מחליף, תרגום מקצועי או שיטות תקשורת אחרות בין מטפלים למטופלים. הפריזמה דרכה מעריכים את ביצועי ה־MT משתנה בהתאם לסקופוס, ובהתאם לכך מתקבלות המלצות שונות.

איכות התרגום

מטרת השימוש השנייה בתרגום מכונה שתיבחן כעת היא איכות התרגום – האם המשתמשים מחפשים תרגום מדויק ואדקווטי, או תרגום טוב דיו?

במאמר הראשון, עורכי המחקר מדדו את איכות התרגום של GT בשבע שפות שונות לפי מדדים של שטף, אדקווטיות, משמעות וחומרה קלינית. רמת הדיוק קיבלה משקל רב בהערכת התרגום, ובבדיקה כוללת של הוראות השחרור התקבל ציון טוב למדי – הועברה המשמעות המדויקת ב־330 מתוך 400 הוראות, שהם 82%. אולם מכיוון שרמת הדיוק השתנתה מאוד משפה לשפה (55%–94%) המליצו בסופו של דבר החוקרים לא להשתמש בתרגום מכונה עבור תרגום מכתבי שחרור מחדר מיון. החוקרים אף ציינו במגבלות המחקר שייתכן שהוא מעריך יתר על המידה את שיעורי הדיוק ושהוא אינו מייצג את ההבנה של מהגרים חדשים, חד־לשוניים. אין התייחסות במאמר בנוגע למידת המובנות של התרגומים, כלומר אם הם היו טובים דיים ומובנים מספיק למטופלים כדי שיוכלו לבצע את אותן הוראות שקיבלו בשחרור למרות טעויות תרגומיות שונות.

במאמר השני נבדקו תרגומים בע"פ של ביטויים הנאמרים בדרך כלל בשיחות פשוטות בין מטפלים למטופלים לפי מדדים של דיוק (אובדן מידע בתרגום), שטף (דקדוק), משמעות וסיכון קליני (האם שינוי המשמעות בתרגום עלול לגרום לסכנה למטופל). התקבלו ציונים נמוכים שלא עמדו ברף האי־נחיתות ועל כן המליצו עורכי המחקר שלא להשתמש בתרגום מכונה לצורך תרגום בעל פה עבור שיחות בין מטפלים למטופלים. אולם, במגבלות המחקר הם ציינו כי הם לא בדקו אם תרגומי מכונה היו מובנים למטופלים, אם הם היו מסוגלים להסיק את התוכן המקורי גם אם התרגום לוקה בחסר בדיוק המילוני, הדקדוקי, הסגנוני או הבהירות. כלומר ייתכן שהתרגום שהתקבל לא היה מעולה, אך היה טוב וברור דיו. חשוב להתעכב על הנקודה הזו במיוחד בסיטואציה זו, של תרגום שיחות פשוטות בעל פה. אפשר לבדוק יחסית בקלות אם המטופל הבין את כוונת המטפל – "תראה לי איפה כואב", והמטופל יצביע על מקום הכאב, "האם אתה מרגיש בחילה?", והמטופל יהנהן או יניד בראשו. אם התרגום מובן מספיק למטופלים, והוא מקל את מחסומי השפה בין הצדדים ומאפשר תקשורת, אפשר להגיע למסקנות והמלצות שונות.

במאמר השלישי, ההתייחסות לדיוק בתרגום מתייחסת באופן ספציפי לדיוק שפה רפואית, ולא לתרגום מדויק באופן כללי. בהמלצות החוקרות לעבודה עתידית, הן מציעות לשלב NMT עם ביטויים מתורגמים באופן מקצועי כדי לוודא שהחלק הרפואי של המלל תמיד יהיה ללא שגיאות. כמו כן, הן מציעות לשפר תרגום מכונה כך שיתייחס לבדיקת ההבנה של המטופל, לעצב תכונות ספציפיות שיאפשרו ביתר קלות לבדוק אם המסר הובהר ולהציע הזדמנויות להסביר לפי צורך. הן מציינות כי ממצאי המחקר מצביעים על כך שהדבר שהכי חשוב לרופאים הוא לא איכות התרגום לשמה, אלא כיצד כלי תרגום מסוגל לתמוך בתקשורת יעילה יותר בין מטופל למטפל באופן בטוח, ובכך לסייע למטופל ולשמור על בריאותו. כלומר, התרגום באופן כללי לא חייב להיות מושלם וחף מטעויות לשוניות, אלא רק החלק הרפואי של הטקסט חייב להיות ללא טעויות כדי לא לסכן את המטופלים והמסר הכללי אינו צריך להיות ברמת הדיוק הכי גבוהה, אלא ברור ומובן דיו.

במאמר הרביעי, החוקרים ציינו במפורש כי בביצוע ההערכה, הם בחרו להתמקד בהבנת המסר המרכזי ובאיכות הכללית ולהשמיט את מדדי הדיוק והשטף הנפוצים בהערכת תרגום. זאת מכיוון שמדדים אלה אקדמיים מדי למטרותיהם: הם התמקדו בהבנה הכללית של התוכן המתורגם, והיה פחות חשוב שהדקדוק שבתרגומים יהיה מושלם או שהרהיטות תהיה ברמה גבוהה. החיסרון של מחקר זה הוא שאין בו פירוט על בדיקת אמינות התרגום ספציפית בהקשר רפואי. ישנה התייחסות לכך שפיתוח הכלי התרחש בזמן מגפת הקורונה, אך חסרה התייחסות לרמת הסיכון הרפואי של התרגומים השגויים. החוקרים ציינו כי לפחות חלק מהמסר המרכזי של טקסט המקור הועבר ב־75% מהמקרים, ושב־48% מהתרגומים המסר המרכזי הועבר לגמרי, אך לא פירטו לגבי שאר התרגומים. חשוב לבדוק אם התרגומים השגויים שבהם לא הועבר המסר בשלמותו או בחלקו לא הציבו סיכון בריאותי למקבלי התרגום, משתמשי הצ'אטבוט. מידע זה חשוב מאוד בהקשר רפואי בכלל ובהקשר של מגיפה עולמית במדינת עולם שלישי בפרט, במיוחד לאור העובדה שבניגוד למסגרות של המאמרים הקודמים, שבהם ציינו דרכים שונות שבהן אפשר לוודא שהמטופל הבין את התרגום כמו תרגום חוזר או ביצוע פעולה לאחר התרגום, כאן המשתמשים ניצבו מול צ'אטבוט ולא מול איש מקצוע רפואי שיכול לוודא את הבנת המסר.

**סיכום**

בעבודה זו נבחנו ארבעה מאמרים העוסקים בתרגום מכונה בסביבות רפואיות. מבט משותף על ארבעת המאמרים העלה שתי סוגיות מרכזיות. הראשונה עסקה בגישה המקדימה לאופן השימוש בתרגום מכונה – שימוש מושכל ו/או חלקי, תוך התערבות והתאמת הטקסט לפי צרכי המשתמש על מנת להפיק תרגומים איכותיים, לעומת גישה המעריכה את איכות תרגום המכונה כפי שהוא בשלמותו, ללא כל התערבות והתאמה מצד המשתמש.הסוגיה השנייה עסקה בהערכת איכות תרגום מכונה לאור תיאוריית הסקופוס, אם מטרת השימוש בתרגום מכונה בסביבות רפואיות היא השגת תרגומים מדויקים לכל אורך הטקסט וחלקיו השונים, או השגת תרגומים טובים דיים, תוך התייחסות להקשר הרפואי הכולל דאגה לשלום המטופלים ולבריאותם.

חשוב להסתכל על תרגום מכונה ככלי שיכול לסייע רבות אם משתמשים בו נכון. כיום, אי אפשר להסתמך עליו באופן בלעדי, אבל אם מזינים טקסט שיהיה קל יותר לתרגום – סימני פיסוק נכונים, שפה פשוטה, תיקון לפי הצורך וכדומה – אפשר לקבל פלט שהסיכויים שלו להיות נכונים הם הרבה יותר גבוהים וכך יהיה אפשר להיעזר בכלי הזה בסביבות שונות בכלל ובמסגרות רפואיות בפרט. הדרכה לצוותים רפואיים לגבי שימוש נכון בכלי תרגום מכונה תוכל להקל את עומס העבודה בתקשורת מול מטופלים שאינם דוברים את שפת הצוות ולסייע להם להעניק טיפול טוב יותר לחולים שזקוקים לתרגום. עם זאת, חשוב לזכור ולשים לב מהי איכות התרגום הרצויה והדרושה עבור כל סיטואציה – אם אפשר להסתפק בתרגום טוב דיו שיהיה מובן למטופל עבור היבטים פשוטים בטיפול, או שהתרגום חייב להיות מדויק לחלוטין, במצבים של סיכון גבוה, ואז חייבים להיעזר בתרגום מקצועי.

ניתן לבדוק במחקרים עתידיים מה איכות התרגום שמתקבלת לאחר התערבויות פשוטות וקלות לביצוע מצד המשתמשים. אם רמת הדיוק של התרגום תעלה באופן ניכר עם התערבות קלה, אפשר יהיה להטמיע את השימוש בה באופן נרחב במסגרות רפואיות שונות במצבים בעלי סיכון נמוך. כמו כן, אם יתקבלו רמות דיוק שונות עבור שפות שונות, אפשר לכתוב הנחיות ברורות הכוללות לא רק כיצד יש להשתמש בתרגום מכונה, אלא גם באילו שפות יש להשתמש בכלי.

**ביבליוגרפיה**

Dew, K. N., Turner, A. M., Choi, Y. K., Bosold, A., & Kirchhoff, K. (2018). Development of machine translation technology for assisting health communication: A systematic review. *Journal of biomedical informatics*, *85*, 56-67.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv preprint arXiv:2004.09095*.

Lee, W., Khoong, E. C., Zeng, B., Rios-Fetchko, F., Ma, Y., Liu, K., & Fernandez, A. (2023). Evaluation of Commercially Available Machine Interpretation Applications for Simple Clinical Communication. *Journal of General Internal Medicine*, 1-7.

Lunt, N., Smith, R. D., Exworthy, M., Green, S. T., Horsfall, D. G., & Mannion, R. (2011). Medical tourism: treatments, markets and health system implications: a scoping review.

Mehandru, N., Robertson, S., & Salehi, N. (2022, June). Reliable and safe use of machine translation in medical settings. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2016-2025).

Öktem, A., DeLuca, E., Bashizi, R., Paquin, E., & Tang, G. (2021). Congolese swahili machine translation for humanitarian response. *arXiv preprint arXiv:2103.10734*.

Taira, B. R., Kreger, V., Orue, A., & Diamond, L. C. (2021). A pragmatic assessment of google translate for emergency department instructions. *Journal of General Internal Medicine*, *36*(11), 3361-3365.

Vermeer, Hans J., 2000. "Skopos and commission in translational action", in *The Translation Studies Reader*, ed. Lawrence Venuti, London and New York: Routledge, pp. 221-233.