אוניברסיטת בר אילן

המחלקה לתרגום וחקר התרגום

הצעת מחקר לתואר השני

**הערכת איכות של תרגום מכונה**

**באמצעות שיטה רב ממדית לניתוח שגיאות (MQM):**

**הנגשת מידע במצבי חירום**

Machine Translation Quality Assessment by means of the Error Analysis Framework of Multidimensional Quality Metrics (MQM):

Information Accessibility in Emergency Situations

**בהנחייתו של ד"ר עמרי אשר**

**יוליה לויט 306094582**

**‏‏‏‏‏‏י"ט סיון תשפ"ב ‏‏18 יוני 2022**

תוכן העניינים

1. נושא המחקר ומטרותיו..................................................................................................2

2. סקירת ספרות ........................................................................................................ 3

2.1 תשתית תאורטית-מושגית כללית ............................................................... 3

2.2 שימוש בשיטות של ניתוח טעויות במחקר..................................................... 7

3. שיטת המחקר ........................................................................................................ 9

4. שאלות המחקר........................................................................................................10

5. השערות..................................................................................................................11

6. תרומת המחקר........................................................................................................12

7. ביבליוגרפיה ........................................................................................................... 12

**1. נושא המחקר ומטרותיו**

בעשורים האחרונים גובר השימוש בתרגום מכונה (Machine Translation ̶ MT) בהקשרים וברבדים חברתיים שונים במדינות רבות בעולם. כלי נגיש זה מאפשר תקשורת בין דוברי שפות שונות, במיוחד כאשר מדובר בנושאים יום-יומיים פשוטים.

במחקר זה בכוונתי להעריך את איכותו של תרגום מכונה בצמדי השפות אנגלית-עברית ואנגלית-רוסית, באמצעות שיטה של ניתוח שגיאות (Error Analysis). מטרת המחקר העיקרית היא לבחון את יעילותו של תרגום מכונה – באמצעות האפליקציה הנפוצה ביותר שלו, Google Translate – בהפקת תרגום סביר ומובן של טקסטים העוסקים במצבי חירום. אתמקד בטקסטים שהבנתם חיונית לביטחונם האישי ולשרידותם של הקוראים במצבים קשים כגון מגפות או מלחמות.

על אף התפתחותו המהירה של תחום תרגום המכונה, מקובל להניח שתוצריהן של מערכות תרגום מכונה עדיין זקוקים לפעולות עריכה המבוצעות בידי מתרגמים ועורכים אנושיים (Post-Editing). ספרות המחקר העוסקת בזיהוי ובסיווג שגיאות בפלט של תרגום מכונה מבקשת להעריך את איכות התוצאות של מערכות תרגום מכונה, לזהות את הסוגים השכיחים ביותר של שגיאות, הן לאחר תרגום מכונה הן בשלב השני, אחרי העריכה האנושית, וכן לסמן כיווני פעולה לשיפור מערכות אלה ולצמצום כמות השגיאות וסוגיהן בפלטי תרגום מכונה (Comparin and Mendes, 2017).

במחקר זה אשתמש בשיטה רב ממדית לניתוח שגיאות (Multidimensional Quality Metrics – MQM), אשר הוצגה על ידי ארל לומל (Lommel) ב-2015 במסגרת הפרויקט Quality Translation 21, בחסות האיחוד האירופי. לפי שיטת MQM כל שגיאה בטקסט היעד מסווגת לפי קטגוריות שונות (שיפורטו בהמשך). נוסף על כך, כל שגיאה נבחנת לפי חומרתה. דרגת החומרה קובעת את גובה הקנסות (penalties), ואלה משפיעים על ציון ההערכה הכללי של פלט התרגום. כלומר, הניקוד הניתן לכל שגיאה מובא בחשבון בחישוב איכות התרגום של הפלט כולו, על פי MQM (Lommel, 2013). שיטת MQM תסייע לי במחקר זה בהסקת מסקנות על איכות פלטי תרגום מכונה של טקסטים העוסקים במידע חיוני בנושאי בריאות וביטחון. להערכה זו יצטרפו גם רשמיהם של מתרגמים ועורכים שישתתפו במחקר, בשלב ה"עריכה שאחרי".

להערכת האיכות של פלטי תרגום מכונה של טקסטים חיוניים במצבי חירום נודעת חשיבות מרובה, מכיוון שאיכות טובה של תרגום עשויה לסייע לגורמים המטפלים הנותנים מענה הומניטרי או בריאותי לפליטים, לחולים או לנפגעי אסון, באספקת מידע מהימן ומהיר בעת הצורך.

**2. סקירת הספרות**

**2.1 תשתית תאורטית-מושגית כללית**

במקביל להתפתחות הטכנולוגית העולמית ולתהליכי הגלובליזציה המואצים, הולך וגובר הצורך בתרגומים. מצב זה מהווה קרקע פורייה ליצירת כלים שיספקו גישה רב-לשונית למידע עבור קהלים שונים וישלימו את התרגומים האנושיים, וביניהם תרגום המכונה.

תוצאות המחקר העדכני בתחום תרגום המכונה מראות, כי תרגום מכונה המבוסס על רשתות עצביות מלאכותיות ( NMT̶ Neural Machine Translation ) מהווה כיום פרדיגמה דומיננטית בזירת תרגום המכונה (Rivera-Trigueros, 2021). מערכת התרגום Google Translate היא הנפוצה ביותר מכל מערכות תרגום המכונה, והשימוש בה גובר מיום ליום בתחומים שונים2021) Taira et al.,).

אחת הסוגיות הבוערות ביותר ביחס לתרגום מכונה היא איכות תוצר התרגום שהוא מפיק. איכות זו נתפסת כיום כנחותה מזו של מתרגמים אנושיים מקצועיים. ההנחה המקובלת במחקר היא, שלצורך שיפור האיכות של תרגום מכונה נחוצות הן הערכות אוטומטיות הן הערכות אנושית. כמו כן, להבטחת האיכות של התוצר התרגומי קיים צורך בעריכה שלאחר התרגום, הנעשית בידי אדם (Rivera-Trigueros, 2021). הסבר וסיווג של שגיאות תרגום מהווים בסיס לשיפור של מערכות תרגום מכונה. סטנדרטיזציה והסדרה של תהליך סיווג השגיאות יתרמו לעיצוב הקריטריונים של הערכה אנושית, ובכך יגבילו, לפחות בצורה חלקית, את השפעת המרכיב הסובייקטיבי שבהערכה ויסייעו לייעול תהליך העריכה שלאחר התרגום.

לפי ג'וליאן האוס (House, 2014), ניתוח הערכת תרגום מכונה עשוי לנוע בין שתי גישות: ניתוח ברמת המאקרו, העוסק בסוגיות אידאולוגיות, תפקודיות, מגדריות או משלביות; וניתוח ברמת המיקרו, העוסק בסוגיות של קולוקציה והגדרת יחידות לשון נפרדות. אולם יש להביא בחשבון, כי מעריכים שונים נוקטים גישות נבדלות בהקשרים שונים.

ככלל, הערכת איכות בתעשייה מתרכזת לרוב בתוצר הסופי או בלקוח, ואילו בתחום חקר התרגום, כאשר אנו עוסקים בתרגום מכונה, הדגש מושם על שיפור המערכות באמצעות מיפוי הפגמים וחיפוש פתרונות לתיקונם, ולאו דווקא על האיכות האבסולוטית של התוצר התרגומי. ניתן לקבוע הבחנות כלליות גם בין הערכה אנושית (ידנית) לבין הערכה אוטומטית. יתרה מכך, ישנן דרכים אחרות להערכת איכות, ששמות דגש על תהליך עבודת התרגום האנושי ולא על הפלט התרגומי כשלעצמו, לדוגמה באמצעות מדידת המאמץ הקוגניטיבי והקושי הטכני הנדרשים במהלך העריכה שלאחר תרגום המכונה.

ככלל, שיטות הערכה אוטומטיות משוות בין הפלט של מערכת תרגום מכונה לבין תרגום או תרגומים אנושיים מקבילים המשמשים כאמת-מידה. אחת השיטות למדידה מסוג זה מודדת את שיעור המילים השגויות (Word Error Rate – WER) ומתבססת על המושג "מרחק עריכה", המתייחס למידת השוֹנוּת בין שתי מחרוזות תווים, דהיינו, בין מילים או ביטויים בתרגום המוצע על ידי המכונה ובתרגום האנושי (Schepens et al., 2012). בשיטתWER ניתוח מרחק העריכה מתבצע ברמת המילים ולא ברמת הפונמות או האותיות. לדוגמה, מרחק העריכה בין הביטוי "ילדים הלכו אתמול להצגה" לבין הביטוי "ילדים ביקרו אתמול בתערוכה" שווה ל-2, כי נדרשות שתי פעולות עריכה כדי להגיע מן הביטוי הראשון לשני: הלכו => ביקרו, להצגה => בתערוכה. שיטה זו אינה מביאה בחשבון את האפשרות לארגן מחדש את סדר המילים במשפט באופן שלא יפגע באיכות התרגום. במסגרת שיטת WER, החלפות, מחיקות והוספות שקולות זו לזו ושוות ל-1. את מספר פעולות העריכה הנחוצות מחלקים במספר המילים שבתרגום האנושי, והאיכות של תרגום המכונה של המשפט מדורגת בהתאם ליחס המתקבל בחלוקה זו. החיסרון העיקרי בשיטת WER נובע מהעובדה שכל מילה בפלט התרגום מושווית למילה הנמצאת *באותו המיקום* בטקסט המקביל של התרגום האנושי. שיטות מדידה מאוחרות יותר, לדוגמה: Translation Error Rate – TER ו- Position-Independent Word Error Rate – PER, ניסו להציע פתרון לבעיה זו. כך, PER מזהה מילים דומות ומשווה ביניהן ללא התייחסות לסדר המילים במשפטים בפלט של תרגום המכונה ובתרגום האנושי המקביל. שיטת המדידה TER מחשיבה ארגון מחדש של מילים כפעולת עריכה נוספת הגוררת קנס נפרד אשר מובא בחשבון בחישוב הניקוד הכללי (Rivera-Trigueros, 2021).

שיטת המדידה האוטומטית הנפוצה ביותר כיום היא שיטת ההערכה הדו לשונית (Bilingual Evaluation Study – BLEU), המודדת דמיון בין פלט המכונה לבין התרגום האנושי ברמה של n-grams, כלומר של יחידות לשון בגדלים שונים – פונמות, הברות, אותיות, מילים או ביטויים, בהתאם לצורך (Papineni et al., 2002). לפי העיקרון המנחה של השיטה, ככל שיש יותר אלמנטים משותפים לפלט המכונה ולתרגום האנושי המקביל, כך מתקבל ניקוד גבוה יותר, והתרגום נחשב לאיכותי יותר. במסגרת השיטה, תרגום קצר מדי יגרור קנס – brevity penalty – מתוך הנחה שהתרגום הקצר מעיד על השמטת מילים ומידע בתהליך תרגום המכונה. למשל, אם רצף המילים "החתול התיישב על הרצפה" בתרגום האנושי מופיע בתור "החתול התיישב" בפלט תרגום מכונה, יופעל קנס על תרגום קצר מדי (Bouillon, 2012).

אף שהשימוש בשיטות דוגמת BLEU נפוץ בתעשייה ובמחקר, חוקרים לא מעטים מצביעים על אי דיוק של השיטות האוטומטיות מסוג זה בהערכת איכות התרגום בפלטי תרגום מכונה, ובמיוחד במערכות המבוססות על רשתות עצביות מלאכותיות הרווחות כיום (Shterionov et al., 2018).

בשונה מהשיטות האוטומטיות, בשיטות הערכה אנושיות העורך או המתרגם האנושי מזהה שגיאות בפלטי תרגום וממיין ומנתח אותן על בסיס קריטריונים שנקבעו מראש (Stymne and Ahrenberg, 2012). בספרות המחקר המוקדמת בתחום, מיון השגיאות בפלטי התרגום התבסס על סיווגי שגיאות שהוצגו במאמריהם של ליטיוס ועמיתיו (Llitjós et al., 2005), וילאר ועמיתיו (Vilar et al., 2006) ופופוביץ ובורשרד (Popovic and Burchardt, 2011). הסיווגים המתוארים במאמרים אלה סכמטיים יותר ממיון השגיאות של שיטת MQM (שתוצג להלן), אך בהחלט ניתן להבחין בקווי דמיון בין שני הסיווגים. ככלל, כיוון שמטרותיה של הערכת איכות התרגום עשויות להיות שונות, ניתן להתאים מערכים של סיווגי שגיאות למטרת המחקר או ליעד התעשייתי המבוקש.

השיטה הרב ממדית להערכת איכות ([Multidimensional Quality Metrics](https://www.w3.org/community/mqmcg/) – MQM) פותחה, כאמור, ב-2015. המודל פותח במסגרת הפרויקט המחקרי QTLaunchPad והוא מתבסס על המודל LISA QA v3.1 (Localization Industry Standards Association), שהושק בשנות התשעים של המאה הקודמת. מודל LISA, בדומה ל-MQM, קבע דירוג של טעויות לפי חומרתן: מינורית, משמעותית וקריטית, ושִׁיֵּךְ לכל רמת חומרה ניקוד מסוים: 1, 5 ו-10, בהתאמה. במובן זה, מודל LISA הוא האבטיפוס של מודל ה-MQM.

לפי הגדרת יוצרי מודל MQM, המודל מספק מסגרת לשיטת מדידה המתארת ומגדירה את איכות התרגום. הוא משמש לא רק כדי לחשב את איכות התרגום על סמך כמות הטעויות וחומרתן אלא גם כדי לסווג את הטעויות לפי מגוון רחב של קטגוריות (שיפורטו בהמשך). מודל MQM מפורט ומעודכן יותר מהשיטות שקדמו לו, שהציגו סיווג מצומצם יותר של טעויות. לדוגמה, במודל של ליטיוס היו חמש קטגוריות עיקריות של טעויות אפשריות: מילה חסרה (missing word), מילה עודפת (extra word), סדר מילים שגוי (wrong word order), מילה שגויה (incorrect word) ואי-התאמה תחבירית (wrong agreement) (Llitjós et al., 2005, p.89). במודל של וילאר, המבוסס על זה של ליטיוס, הושמטו הקטגוריות של מילה עודפת ואי-התאמה תחבירית, אך נוספו קטגוריות של מילים לא ידועות (unknown words) ופיסוק (punctuation) (Vilar et al., 2006, p.699).

שש הקטגוריות העיקריות למיון שגיאות בשיטת MQM הן: "דיוק" (Accuracy), "רהיטות" (Fluency), "אמיתוּת" (Verity), "מינוח" (Terminology), "סגנון" (Style), "מוסכמות מקומיות" (Locale convention) ו"אחר" (Other).

תחת הקטגוריה "דיוק" נמצא שגיאות הנובעות מחוסר דיוק בהמרת טקסט המקור לטקסט היעד, כגון: הוספה (Addition), תרגום לא מדויק או לא נכון (Mistranslation), השמטה (Omission) וחלקים לא מתורגמים (Untranslated). בקטגוריה "רהיטות" נכללות שגיאות הנובעות מצורתו או מתוכנו של הטקסט. סוגי השגיאות בקטגוריה זו מתייחסים לתחום הדקדוק (Grammar), לחוסר עקביות בטקסט (Inconsistency), לאיות (Spelling), למבע בלתי ניתן להבנה (Unintelligible), לכפילות (Duplication), למשלב לא מתאים (Register) ולטיפוגרפיה (Typography).

תחת הקטגוריה "מוסכמות מקומיות", נמצא שגיאות הנובעות מאי התאמה למוסכמות של שפת היעד, כולל אי התאמות הנובעות מהבדלים תרבותיים בין תרבות המקור לתרבות היעד. בקטגוריה "מינוח" מנותחות שגיאות הנובעות מתרגום אחר מהמצופה למונחים בתחום הספציפי, ותחת כותרת זו נכללות גם שגיאות של חוסר עקביות במינוח (Term Inconsistency).

הקטגוריה "אמיתוּת" עוסקת במצבים שבהם הצהרות בטקסט מנוגדות להיגיון הפנימי שלו, לדוגמה, עובדות שהן אמת בסביבת המקור, אך אינן תואמות את המציאות בסביבת היעד. שגיאות מסוג זה עשויות להיווצר מחוסר שלמות (Completeness) של המבע, כשתוכן הכרחי חסר בטקסט המקור. למשל, אם טקסט המקור עוסק בתיאור הרכבת מכשיר כלשהו, ובתיאור חסרים שלבים חיוניים להשלמת התהליך, מתקבל תרגום חסר, והדבר עלול לפגוע בתקינות המכשיר ובאופן פעולתו. יש להבדיל בין הסעיף "חוסר שלמות" לבין סעיף ההשמטה (Omission), המאופיין בהשמטה של חלק מטקסט המקור בטקסט היעד. דוגמה נוספת לשגיאות בקטגוריה "אמיתות" היא תוכן ספציפי לאזור מסוים (Locale Specific Content), כלומר, תוכן מיוחד ל"לוקאל" מסוים שאינו מתאים לקהל היעד, ל"לוקאל" או למטרת הטקסט. במקרים כאלה קיימת התאמה למוסכמות מקומיות (לדוגמה, התאמת סוג המטבע במדינות שונות), אך ישנו פער בין המציאות התרבותית של שפת המקור ושל שפת היעד. למשל, טקסט המתייחס להודעה פרסומית במדינה מסוימת (בשפת המקור) עלול להיות בלתי רלוונטי לקהל היעד במדינה אחרת (בשפת היעד).

חשוב להדגיש כי הסעיף "תוכן ספציפי לאזור מסוים", הנמצא בקטגוריה "אמיתות", שונה מקטגוריית "מוסכמות מקומיות" (Locale convention), העוסקת בעיקר במאפיינים הפורמליים, כגון, המטבע או יחידות המדידה הנהוגות באזורים שונים.

תחת הקטגוריה "סגנון" נכללות שגיאות הנובעות מן הבעיות הסגנוניות של הטקסט. ולבסוף, הקטגוריה "אחר" מכילה את כל הטעיות שלא ניתן למיין אותן בקטגוריות הקיימות ב-MQM.

במסגרת MQM, יש לייחס לכל שגיאה רמת חומרה מסוימת, שלפיה נקבע גובה הקנס (penalty). הדרגה הנמוכה ביותר היא "אין טעות" (none): גובה הקנס הוא 0. שגיאות בדרגת חומרה זו הן כאלה שיש לשים אליהן לב, אך קנסות אינם מוטלים עליהן. בקטגוריה זו נמצאות "טעויות העדפה" (preferential errors), כלומר מקרים שבהם המבע המתורגם אינו שגוי, אך ככל הנראה מתרגם אנושי היה מוצא להם פתרון תרגומי אחר, מדויק ואלגנטי יותר. כמו כן, בקטגוריה הזו נכללות טעויות שחוזרות על עצמן בשיטתיות ושניתן לתקנן בקלות, לדוגמה, שימוש שיטתי במושג שגוי שקל להחליפו בפעולה פשוטה של חיפוש והחלפה. דרגת החומרה הבאה היא "טעות מינורית" (minor): גובה הקנס הוא 1. שגיאות מינוריות הן שגיאות אשר אינן משפיעות על שימושיות (usability) ועל יכולת ההבנה (understandability) של התוכן. כלומר, אם קורא טיפוסי מסוגל לתקן את הטעות באופן אמין בלי שהטעות תשפיע על שימושיות התוכן, דרגת החומרה של הטעות היא מינורית. הדרגה השלישית היא "טעות משמעותית" (major): גובה הקנס הוא 10. טעויות מסוג זה הן טעויות המשפיעות על השימושיות ועל יכולת ההבנה של התוכן, אך בו בזמן אינן הופכות את הטקסט לבלתי שמיש. איות שגוי של מילה, לדוגמה, ידרוש מהקורא זמן ומאמץ רב יותר כדי להבין את משמעות המילה, אך הדבר יעלה בידו. ולבסוף, הדרגה הרביעית היא "טעות קריטית" (critical): גובה הקנס הוא 100. טעויות קריטיות הן טעויות ההופכות את התוכן לבלתי שמיש. לדוגמה, טעות דקדוק חמורה במיוחד המשנה את משמעות הטקסט תיחשב לטעות קריטית. במידה שהטעות מונעת מהקורא שימוש בתוכן כפי שתוכנן או במידה שהטעות מציגה מידע שגוי העלול להסב נזק לקורא או למשתמש, היא תיחשב לטעות קריטית.

במאמר משנת 2013 מציגים לומל ועמיתיו (Arle Richard Lommel et al.) את הנוסחה לחישוב האיכות לפי שיטת הניקוד של MQM:

***TQ = 100 − AP − (FPT − FPS) − (VPT − VPS)***

בנוסחה זו: **TQ** – ניקוד האיכות הכולל (Total Quality Score), הוא התוצאה הסופית שמביאה בחשבון את כל הגורמים הרלוונטיים לקביעת רמת איכות התרגום לפי שיטת MQM; **AP** – קנסות המוטלים על טעויות הדיוק (סכום של כל הקנסות בקטגוריה של דיוק – Accuracy); **FP***T* – קנסות המוטלים על טעויות ברהיטות של טקסט היעד (סכום הקנסות בקטגוריה של רהיטות – Fluency – בשפת היעד); **FP**S – קנסות המוטלים על שגיאות ברהיטות של טקסט המקור, כלומר, סכום הקנסות בקטגוריה של רהיטות – Fluency – בשפת המקור (אם אין התייחסות להערכת האיכות של טקסט המקור **FP***S* =0*);* **VP***T* – קנסות המוטלים על שגיאות באמיתות של טקסט היעד, כלומר, סכום הקנסות בקטגוריה של אמיתות – Verity – בשפת היעד; ו- **VP**S – קנסות המוטלים על שגיאות באמיתות של טקסט המקור, כלומר, סכום הקנסות בקטגוריה של אמיתות – Verity – בשפת המקור (אם אין התייחסות להערכת האיכות של טקסט המקור **VP***S*=0*).*

חישוב הקנסות מתבצע לפי הנוסחה הבאה, שגובה הקנסות בה עודכן לפי המוסכמות החדשות של MQM משנת 2015:

***P = (Issuesminor + Issuesmajor × 10+ Issuescritical × 100) ÷ Word count***

כלומר, סכום הקנסות לקטגוריה מחושב לפי המפתח הבא: סכום המשקלים של כל הטעויות המינוריות, המשמעותיות והקריטיות, כאשר מספר הטעויות מוכפל בגובה הקנסות המוטלים על כל אחת מהטעויות – 1, 10 ו-100 בהתאמה – מחולק בכמות המילים בטקסט.

חוקרים רבים מציינים כי יש יתרונות וחסרונות הן לשיטות האוטומטיות הן לשיטות האנושיות להערכת איכות תרגום מכונה (ראו, למשל: Rivera-Trigueros, 2021). בשיטות האוטומטיות העלות אומנם נמוכה יותר, המאמץ האנושי קטן יותר ורמת האובייקטיביות גבוהה יותר, אך הן תלויות בתרגומים אנושיים מקבילים והיכולת שלהן להעריך שקילוּת (equivalence) סמנטית ותחבירית מוגבלת מאוד. בניגוד לכך, שיטות אנושיות להערכת תרגום מכונה נחשבות לאמינות יותר, והביקורת העיקרית כלפיהן מתמקדת בכך שהן דורשות משאבים רבים, לרבות משאבי אנוש, השקעת זמן, הכשרת המעריכים, וכולי. במחקר זה בחרתי להשתמש בשיטת סיווג השגיאות האנושית MQM, זאת מכיוון שמטרת המחקר היא להגיע להערכה המדויקת ביותר של איכות הפלט התרגומי. היות שקנה המידה של המחקר שלי מצומצם לאין שיעור לעומת קנה המידה הנדרש בתעשייה, סוגיית המשאבים בהקשר זה אינה משמעותית.

**2.2 שימוש בשיטות של ניתוח טעויות במחקר**

בשנים האחרונות התפרסמו מקרי בוחן (case studies) שונים בתחום ניתוח הטעויות (Error Analysis), שעשו שימוש בשיטות ניתוח שונות בהתאם ליעדי המחקר. שניים מן המחקרים האלה – של מסארו יאמדה (Yamada, 2019) ושל שרה מנדס ולוסיה קומפרין (Comparin and Mendes, 2017) – ישמשו בסיס לעבודתי.

יאמדה בחן באמצעות מערכת MNH-TT (Minna no Hon’yaku for Translator Training) – פלטפורמה שיתופית להכשרת מתרגמים המבוססת על מערכת זיהוי וסיווג שגיאות – את כשירותם של סטודנטים לאנגלית ביפן לערוך פלט של תרגום מכונה ולזהות את השגיאות בתרגום (Yamada, 2019). השיטה לסיווג טעויות ששימשה את יאמדה במחקרו כוללת תפריט מוכן של קטגוריות מיון השגיאות (Toyoshima, et al. 2016; Yamamoto et al., 2016). להבדיל משיטת MQM, שבמסגרתה העורך או המתרגם בוחר בין קטגוריות שונות של סוגי שגיאות, פלטפורמת MNH-TT מציגה עצי החלטה (decision trees) המכוונים את הסטודנטים בתהליך קבלת החלטות במהלך העריכה של פלט תרגום מכונה. כלומר, פלטפורמת MNH-TT מכוונת את אלה המשתמשים בה בתהליך קבלת החלטות, ואילו שיטת MQM מאפשרת לעורכי התרגום לבחור באופן עצמאי מקטלוג מגוון של סוגי שגיאות. בדומה למחקרו של יאמדה, גם במחקר שלי ימוינו כל השגיאות בפלט תרגום המכונה לפי קטגוריות שונות של שיטה לניתוח שגיאות, ולכל שגיאה תיוחס דרגת חומרה מסוימת. כמו כן, בדומה ליאמדה, ששילב במחקרו מידע שהתקבל משאלונים שמילאו משתתפי המחקר, בכוונתי להעריך את התרשמותם האישית של המשתתפים מכמות וממורכבות "העריכה שאחרי", באמצעות שאלונים.

מטרת המחקר של שרה מנדס ולוסיה קומפרין (Comparin and Mendes, 2017), לעומת זאת, הייתה כפולה: בשלב הראשון ערכו החוקרות מיון וניתוח של הסוגים הבולטים ביותר של שגיאות בפלט של תרגום מכונה מאנגלית לאיטלקית, ובשלב השני בחנו את השכיחות של סוגי השגיאות השונים שנותרו בטקסט לאחר עריכה אנושית. מנדס וקומפרין מדווחות על צמצום משמעותי במספר המוחלט של שגיאות בשלב השני לעומת השלב הראשון. עם זאת, עולה מן המחקר כי ישנם סוגים מסוימים של שגיאות שנמצאו בפלט גם אחרי העריכה האנושית. לטענת החוקרות, על בסיס כמות השגיאות בפלט של תרגום המכונה ושכיחותם של סוגים מסוימים של שגיאות, ניתן להצביע על כיוונים לפיתוחים עתידיים לשיפור מערכות תרגום מכונה, לנטרול פגמיהן ולהקטנת מאמץ תרגומי בעריכה אנושית שלאחר תרגום מכונה. בדומה למנדס וקומפרין, אני מתכוונת לזהות ולמיין את השגיאות בפלט תרגום המכונה תוך כדי עבודת עריכה ראשונית שלי על הטקסט המתורגם. בשלב השני, אשלח את פלטי תרגום המכונה המקוריים (ללא העריכה שלי) לעורכים אחרים, ואלה יבצעו בעצמם "עריכה שאחרי" ומיון שגיאות. כמו מנדס וקומפרין, בכוונתי להשוות בין כמות השגיאות שתוקנו בידי עורכים אלה לבין כמות השגיאות שלא תוקנו, ולנתח את התפלגות השגיאות לפי הקטגוריות.

חשוב להדגיש כי המיזם התרגומי העומד במרכז המחקר של מנדס וקומפרין הוא הפלטפורמה לתרגומים Unbabel. פלטפורמה זו מגייסת עובדים על בסיס מיקור המונים (Crowdsourcing) ומקבלת לשורותיה כל אדם דו-לשוני שעובר בהצלחה את מבחן ההתאמה, גם אם הוא חסר ניסיון והכשרה בתחום התרגום, ומציעה לו שכר עבור שירותי העריכה שיספק. לאור פרופיל העורכים האנושיים שהשתתפו במחקרן של מנדס וקומפרין, וכן במחקרו של יאמדה, נשאלת השאלה מהי רמת מיומנותם של משתתפי המחקר בתור מתרגמים ועורכים. במחקרו של יאמדה שאלה זו נוסחה כאחת משאלות המחקר – יאמדה השתמש בשיטת סיווג שגיאות כדי לקבוע את רמת המיומנות של הסטודנטים בתחום התרגום ועריכת התרגום, וציין במפורש כי הסטודנטים אינם מתרגמים מקצועיים. ג'נס דריימיקר, שעסק בהערכת איכות של תרגומים שיתופיים בתחום של ספרי תיירות, השווה בין ספרי תיירות שתורגמו לספרדית בידי מתרגמים מקצועיים לבין תרגומים ניסיוניים במסגרת פלטפורמת Duolingo המבוססת על מיקור המונים (Deriemaeker, 2014). תוצאות ההשוואה מעניינות: הכמות הכוללת של שגיאות בתרגומים מקצועיים משתווה לכמות הכוללת של שגיאות בתרגומים לא מקצועיים, אך בבחינה מדוקדקת יותר ניתן להבחין כי דרגת חומרת השגיאות בתוצריהם של מתרגמי Duolingo גבוהה יותר מדרגת חומרתן בתרגומים המקצועיים. מכאן ברור כי האיכות הכללית של תרגומים מקצועיים גבוהה יותר מאיכות התרגומים המופקים באמצעות מיקור המונים.

במחקרי זה, בניגוד למחקרו של יאמדה, אין בכוונתי להעריך את מיומנותם של משתתפי המחקר העורכים את פלטי תרגום מכונה. יתרה מכך, אני שואפת לצמצם במידת האפשר את כמות שגיאות העריכה ואת מקרי אי האיתור של שגיאות בפלט תרגום המכונה. לפיכך, יגויסו למחקר זה סטודנטים מתקדמים לתרגום ומתרגמים מקצועיים בלבד.

בדומה למחקרן של מנדס וקומפרין, אני מתכננת לבדוק הן את כמות השגיאות הן את שכיחות הסוגים השונים של השגיאות בפלט תרגום מכונה. עם זאת, אוסיף למחקר את ההשוואה בין שני צמדי שפות – אנגלית-רוסית ואנגלית-עברית – שלא הייתה קיימת אצל מנדס וקומפרין. להבדיל מהמחקרים שהזכרתי לעיל, המחקר שלי אף יתמקד בהערכת האיכות של פלטי תרגום המכונה, כשהנחת היסוד היא שמלאכת "העריכה שאחרי" נעשית בידי מתרגמים ועורכים בעלי מיומנויות תרגום גבוהות.

**3. שיטת המחקר**

במסגרת המחקר אבחר שני טקסטים באורך כ-300 מילים. הטקסטים יציגו מידע חיוני שתרגומו הבהיר חשוב לביטחונם ולבריאותם של קוראי הטקסט, לדוגמה, הנחיות בידוד לחולי קורונה או מידע חיוני לפליטים מאזורי הלחימה באוקראינה. הטקסטים יתורגמו באמצעות מערכת Google Translate מאנגלית לעברית ומאנגלית לרוסית. בדומה למנדס וקומפרין, בכוונתי לזהות ולמיין את השגיאות בפלט תרגום המכונה תוך כדי עריכה. בהמשך, אמסור את פלטי תרגום מכונה לא ערוכים לעורכים נוספים כדי שיבצעו "עריכה שאחרי" ומיון שגיאות. לאחר מכן, אערוך השוואה בין כמות השגיאות שתוקנו לבין כמות השגיאות שלא תוקנו ואנתח את התפלגות השגיאות לפי הקטגוריות.

באופן דומה לשיטת המחקר של יאמדה (Yamada, 2019), אני מתכוונת למיין את כל השגיאות בפלט תרגום המכונה לפי קטגוריות שונות של שיטה לניתוח שגיאות, ולייחס לכל שגיאה דרגת חומרה מסוימת; אך בניגוד ליאמדה, אינני שואפת להעריך את מיומנותם של משתתפי המחקר. בדומה למחקרן של מנדס וקומפרין, אבדוק את כמות השגיאות ואת שכיחות הסוגים השונים של השגיאות בפלט תרגום מכונה. כמו כן, אוסיף למחקר את ההשוואה בין שני צמדי שפות – אנגלית-רוסית ואנגלית-עברית – שלא הוצגה במקרה בוחן של מנדס וקומפרין.

כיוון שפלטי תרגום המכונה מתורגמים הן מאנגלית לעברית הן מאנגלית לרוסית, אוכל להשוות את איכות התרגום בין שני צמדי השפות על סמך הניקוד לפי שיטת MQM. ההשוואה בין שני צמדי השפות הללו מסקרנת גם משום שאנגלית ורוסית הן שפות עתירות משאבים (high-resource languages), ואילו השפה העברית היא שפה דלת משאבים (low-resource language). בשפות עתירות משאבים, וביניהן בראש ובראשונה אנגלית, קיימים מאגרים עצומים של מידע ונתונים ברשת. הכמות האדירה של טקסטים בשפות עתירות משאבים מאפשרת פיתוח מערכות ייחודיות לכל שפה ושפה, כשהמערכות הללו מבוססות על למידת מכונה (machine-learning based systems). בשפות דלות משאבים, לעומת זאת, ישנם מאגרים מצומצמים יותר של מידע ונתונים, או לעיתים היעדר מוחלט של טקסטים ברשת, לכן אין מספיק נתונים לפיתוח מערכות שיתאימו למבנים הייחודים של שפות כאלה. לפיכך, שפות דלות משאבים הן בעיקר שפות עם מספר דוברים מצומצם ושפות בעלות מעמד חברתי-פוליטי נמוך.

כאמור, במחקר זה אשתמש בשיטת ההערכה לניתוח שגיאות MQM (Lommel, 2015), המאפשרת לעצב שיטות מדידה ייעודיות המתאימות לנושאים ספציפיים ולמטרות ייחודיות, ובכך נותנת לחוקר יד חופשית בקביעת הממדים שלפיהם תתבצע ההערכה.

במחקר הנוכחי החלטתי לשלב את כל הקטגוריות העיקריות של שיטת ה-MQM:"דיוק", "רהיטות", "אמיתות", "מינוח", "סגנון", "מוסכמות מקומיות" ו-"אחר". כל אחת מהקטגוריות כוללת סעיפים ותתי-סעיפים המגדירים סוגים שונים של שגיאות בתרגום, וביניהם "הוספה", "תרגום לא מדויק או לא נכון", "השמטה", "דקדוק", "חוסר עקביות בטקסט", "איות", "כפילות", "משלב לא מתאים", "טיפוגרפיה" ו-"חוסר שלמות". כפי שפירטתי בסעיף 2.1, הנוסחה שתשמש אותי בחישוב הניקוד (Arle Richard Lommel et al. 2013) היא:

TQ = 100 − AP − (FPT − FPS) − (VPT − VPS)

הנוסחה לחישוב הקנסות, שגובה הקנסות בה עודכן לפי המוסכמות החדשות של MQM משנת 2015 היא:

P = (Issuesminor + Issuesmajor × 10+ Issuescritical × 100) ÷ Word count

דהיינו, בכל קטגוריה מחשבים את הקנסות לפי המפתח הבא: סכום המשקלים של כל הטעויות המינוריות, המשמעותיות והקריטיות, כאשר מספר הטעויות מוכפל בגובה הקנסות המוטלים על כל אחת מהטעויות – 1, 10 ו-100 בהתאמה – מחולק בכמות המילים בטקסט.

לסיכום, מתווה המחקר יכלול את ארבעת השלבים הבאים: 1) בחירת הטקסטים ותרגומם באמצעות Google Translate; 2) סיווג ראשוני של שגיאות וקביעת חומרתן לפי שיטת ה-MQM; 3) שליחת הקטעים (הטקסטים המקוריים ופלטי תרגום המכונה) למתרגמים ולעורכים מקצועיים לביצוע "העריכה שאחרי" של הפלטים וסיווג שגיאות לפי שיטת ה-MQM; ו-4) ניתוח המידע שהתקבל, השוואתו לסיווג הראשוני של שגיאות (שבוצע על ידִי) והסקת מסקנות על אודות איכות תרגום המכונה של הפלטים הנחקרים.

נוסף על מלאכת העריכה יתבקשו משתתפי המחקר, מתרגמים ועורכים מקצועיים, להשיב על שאלונים ובהן שאלות פתוחות, שיתייחסו לניסיונם בתרגום ולחוויית העריכה הנוכחית. ניתוח המידע שיתקבל מן השאלונים יוסיף ממד איכותני למחקר. ההערכה הכמותנית (חישוב הקנסות לפי הנוסחה) בשילוב עם ההערכה האיכותנית (שאלות פתוחות בשאלונים) צפויות לספק תמונת מצב מקיפה. בסִפרן של גביאלה סאלדנה ושרון אובריין, העוסק במתודולוגיות בחקר התרגום, המחברות ממליצות על השיטה המשולבת הזו, הנקראת טריאנגולציה (triangulation), ובמסגרתה מנותחים נתונים המתקבלים במחקר ומושווים לפי שתי גישות שונות (Saldanha and O’Brien, 2014).

השאלונים שישמשו אותי במחקר מבוססים על השאלונים ששימשו את מאסרו יאמדה במחקרו: לאחר העריכה שביצעו המשתתפים לתוצרי תרגום המכונה, הם השיבו על שאלונים, ובהם התייחסו למאמץ שנדרש מהם לצורך העריכה שלאחר תרגום המכונה (Post-Editing Effort) והביעו את דעתם על אודות חוויית העריכה, כל זאת בתשובות לשאלות פתוחות (Yamada, 2019). במחקר מסוג זה, תשובותיהם של משתתפי המחקר צפויות להיות סובייקטיביות, אך כיוון שהתשובות יסכמו את חוויות העריכה של כל משתתפי המחקר וכן יתאזנו באמצעות ניתוח כמותני (חישוב לפי הנוסחה), ניתן להעריך כי המסקנות הכלליות מן השאלונים יספקו לנו תמונת מצב כללית השואפת לאובייקטיביות ככל האפשר.

**4. שאלות המחקר**

בהתבסס על מטרת המחקר ועל סקירת הספרות, ייבחנו במחקר השאלות הבאות:

1. האם תרגום מכונה באמצעות Google Translate עשוי להיות שימושי במצבי חירום הדורשים הנגשה מהירה של מידע מהימן?
2. האם קיימים הבדלים באיכות תרגום המכונה בצמדי השפות הנחקרים (אנגלית-עברית מול אנגלית-רוסית)?
3. האם סוגים מסוימים של שגיאות נפוצים יותר מאחרים, ואילו הם? הסוגייה תיבחן על סמך התיקונים שיופיעו בפלטי תרגום מכונה הערוכים בידי משתתפי המחקר, ועל סמך מיוני השגיאות שעשו.
4. האם שכיחותם של סוגי השגיאות תלויה בצמד השפות בתרגום?
5. האם קיימת התאמה בהערכת איכות תרגום המכונה בין החישוב שנעשה באמצעות הנוסחה של לומל לבין התרשמותם האישית של המתרגמים ושל עורכי התרגום המשתתפים במחקר, על פי העולה מתשובותיהם לשאלון?
6. האם ישנה הסכמה בקרב המתרגמים ועורכי התרגום המשתתפים במחקר בנוגע לרמת האיכות של תרגום המכונה בפלטים שקיבלו?

**5. השערות**

בהתבסס על סקירת הספרות ועל תוצאות המחקרים בתחום הערכת איכות תרגום מכונה, השערותיי הן:

1. תרגום מכונה של פלטי המחקר יהיה שימושי באופן חלקי, ובמידה כזו או אחרת יזדקק לעריכה של גורם אנושי לאחר מכן.
2. צפויה איכות טובה יותר של תרגום מכונה בצמד השפות אנגלית-רוסית לעומת הצמד אנגלית-עברית, זאת מכיוון שהשפה הרוסית היא שפה עתירת משאבים, ואילו השפה העברית היא דלת משאבים.
3. צפוי כי במהלך ניתוח המידע שיתקבל מפלטי תרגום המכונה הערוכים יימצא כי סוגים מסוימים של שגיאות יופיעו בשכיחות רבה יותר מסוגים אחרים. אחת הסיבות לכך היא העושר המורפולוגי היחסי (למשל: נטיות הפועל לציון גוף, מין ומספר וכדומה) של השפה הרוסית והעברית לעומת השפה האנגלית.
4. צפוי כי שכיחותם של סוגים מסוימים של שגיאות תהיה שונה ברוסית ובעברית לנוכח העובדה כי ברוסית ישנה נטיית שמות עצם, שמות תואר ומספרים (יחסות). כיוון שבאנגלית לא קיימת מערכת דומה, צפוי כי בפלטי תרגום מכונה בצמד אנגלית-רוסית יתגלו שגיאות הנובעות משימוש שגוי ביחסות.
5. צפוי כי תהיה התאמה בהערכת איכות תרגום המכונה בין החישוב שנעשה באמצעות הנוסחה של לומל לבין התרשמותם האישית של המתרגמים ועורכי התרגום המשתתפים במחקר.
6. צפוי כי תהיה הסכמה כללית בקרב המתרגמים ועורכי התרגום המשתתפים במחקר לגבי איכות תרגום המכונה בפלטים שיקבלו לעריכה.

**6. תרומת המחקר**

המחקר יתרכז באיכות תרגום מכונה של טקסטים חיוניים שיש צורך לתרגמם במהירות בעיתות מצוקה ובמצבי חירום. במקרים כאלה, תרגום מהיר וסביר עשוי להיות הכרחי לשמירה על בריאותם ועל ביטחונם של נמעני הטקסט המשתמשים בתרגום, ולעיתים אף חיוני להישרדותם.

ככלל, מחקרים העוסקים בהערכת איכות תרגום מכונה מסמנים את סוגי השגיאות הנפוצים ביותר, ובכך עשויים לתרום לשיפור מערכות תרגום מכונה ולסייע במלאכת העריכה שאחרי קבלת תוצר תרגום המכונה (Comparin and Mendes, 2017; Rivera-Trigueros, 2021). למיטב ידיעתי, מחקרים מסוג זה לא נערכו בצמד השפות אנגלית-עברית, ואילו בצמד השפות אנגלית-רוסית התמקדו המחקרים בהיבט של מאמץ העריכה האנושי שאחרי תרגום המכונה (ראו, למשל: Ziganshina, 2021).

לאחרונה, מתהווה מגמה מחקרית העוסקת בחקר תרגום מכונה במצבי חירום, כגון: מגפות, הגירה המונית, רעידות אדמה וכדומה. ב-2011, במסגרת המיזם ההומניטרי של הרווארד (Harvard Humanitarian Initiative), פורסם דו"ח Disaster Relief 2.0, אשר קרא לחוקרים לבחון את התרומה הפוטנציאלית שעשויה להיות לטכנולוגיה בעיתות משבר (Cadwell et al., 2019). תרגום מכונה מעורר עניין מיוחד בהקשר של מצבי חירום, בהיותו אמצעי נגיש לתקשורת בין דוברי שפות שונות, לכן אנו עדים להיווצרות תחום מחקר נפרד בחקר תרגום המכונה העוסק במצבים כאלה – תרגום מכונה למצבי חירום או משבר (Crisis MT).

אומנם אין לצפות שבעתיד הקרוב (ואף הרחוק) יהווה תרגום המכונה תחליף מושלם למתרגמים אנושיים, אך בהחלט ניתן לשפר ולהיטיב את מערכות התרגום כך שיהיו מסוגלות להפיק תרגומים באיכות טובה בצורה מספקת, המובנים למשתמשים הזקוקים לתרגום מהיר ונגיש.

**7. ביבליוגרפיה**

\*Bouillon, P. (2012) Analysis of Existing Metrics and Proposal for a Task-oriented Metric. In *ACCEPT- Automated Community Content Editing PorTal*.

\*Cadwell, Patrick, O’Brien, Sharon, DeLuca, Eric (2019) ‘More than Tweets:

A Critical Reflection on Developing and Testing Crisis Machine Translation Technology’,

Translation Spaces 8(2), pp. 300–333.

\*Comparin, Lucia, Mendes, Sara (2017) ‘Using Error Annotation to Evaluate Machine Translation and Human Post-Editing in a Business Environment’. In *Proceedings of EAMT* 2017, Prague, vol. 31.

\*Daems, Joke, Macken, Lieve, and Vandepitte, Sonia (2014) ‘On the Origin of Errors: A Fine-grained Analysis of MT and PE Errors and their Relationship’. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation,* LREC 2014, Reykjavik, Iceland, May 26-31, 2014, pp. 62-66.

\*Deriemaeker, Jens (2014) ‘The Power of the Crowd: Assessing Crowd Translation Quality of Tourist Literature’, Master Thesis, Ghent University, Belgium.

\*House, Juliane (2014) ‘Translation: A Multidisciplinary Approach’ in *Translation Quality Assessment: Past and Present*, Palgrave Macmillan, pp. 214-264.

Llitjós, Ariadna Font, Carbonell, Jaime G., Lavie, Alon (2005) ‘A Framework for Interactive and Automatic Refinement of Transfer-based Machine Translation’. In *Proceedings of the10th Annual Conference of the European Association for Machine Translation (EAMT),* Budapest, Hungary, May 2005, pp.87-96.

\*Lommel, Arle Richard, Burchardt, Aljoscha, Uszkoreit, Hans (2013) ‘Multidimensional Quality Metrics: A Flexible System for Assessing Translation Quality’ in *Proceedings of Translating and the Computer,* vol. 35, London, UK. Aslib.

\*Lommel, Arle (2015) Multidimensional Quality Metrics (MQM) Definition. URL <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>.

\*Yamada, Masaru (2019) ‘The impact of Google Neural Machine Translation on Post-editing by Student Translators’, *The Journal of Specialised Translation*, issue 31, pp. 87-106.

\*Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). ‘BLEU: A Method for Automatic Evaluation of Machine Translation’. In *Proceedings of the 40th annual meeting on association for computational linguistics.*

\*Popovic, Maja, Burchardt, Aljoscha (2011) ‘Error Analysis of Machine Translation Output’ in *Proceedings of the 15th Annual Conference of the European Association for Machine Translation*, pp. 265-272.

\*Rivera-Trigueros, Irene (2021) ‘Machine Translation Systems and Quality Assessment: a Systematic Review’ in [*Language Resources and Evaluation*](https://link.springer.com/journal/10579), Springer. Published in open access.

\*Schepens, Job, Dijkstra, Ton, Grootjen, Franc (2012) ‘Distributions of Cognates in Europe as Based on Levenshtein Distance’, *Bilingualism: Language and Cognition* 15 (1), pp. 157–166

\*Shterionov, Dimitar, Superbo, Riccardo, Nagle, Pat, Casanellas, Laura, O’Dowd, Tony, Way, Andy (2018) ‘Human Versus Automatic Quality Evaluation of NMT and PBSMT’, Published online: 8 May 2018 ©*Springer Science+Business Media B.V., part of Springer Nature 2018*

\*Stymne, Sara, Ahrenberg, Lars (2012) ‘On the Practice of Error Analysis for Machine Translation Evaluation’ in *Proceedings of the Eighth International Conference on Language Resources and Evaluation*,LREC 2012, Istanbul, Turkey, May 23-25, 2012,pp. 1785-1790.

\*Saldanha, Gabriela, O’Brien, Sharon (2013) ‘Research Methodologies in Translation Studies’, *St. Jerome Publishing*, pp. 22-23.

\*Taira, Breena R., Diamond, Lisa C., Kreger, Vanessa, Orue, Aristides (2021) ‘A Pragmatic Assessment of Google Translate for Emergency Department Instructions’, Journal of General Internal Medicine*.*

**\***Toyoshima, Chiho, Fujita, Atsushi, Tanabe, Kikuko, Kageura, Kyo, Anthony Hartley (2016) ‘Analysis of Error Patterns of Translation Students Based on Revision Categories’, *Interpreting and Translation Studies 16*, pp. 47-65.

\*Vilar, D., J. Xu, L. D’haro, and H. Ney (2006) ‘Error Analysis of Statistical Machine Translation Output’ in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*, Genoa, Italy, May 2006. European Language Resources Association (ELRA), pp. 697-702.

\*Yamamoto, Mayuka, Tanabe, Kikuko, Atsushi Fujita (2016) ‘Changes in Translation Learner's Error Types over Time (Hon'yaku Gakushu-sha noGakushu-Katei ni okeru Error no Keikou no Henka)’. In *Proceedings of NLP2016 (22),* pp. 865-868*.* http://www.anlp.jp/proceedings/annual\_meeting/2016/pdf\_dir/E5-3.pdf (consulted 21.12.2017).

\*Ziganshina, Liliya E., Yudina, Ekaterina V., Gabdrakhmanov, Azat I., Ried, Juliane (2021) ‘Assessing Human Post-Editing Efforts to Compare the Performance of Three Machine Translation Engines for English to Russian Translation of Cochrane Plain Language Health Information: Results of a Randomised Comparison’, Informatics , Vol. 8, Issue 1, p. 9, MDPI.