Kaunas University of Technology

Faculty of Social Sciences, Arts and Humanities

Human Performance in Machine Translation Post-Editing

Master's Final Degree Project

Sandra Vasiliauskienė

Project author

Prof. dr. Ramunė Kasperė

Supervisor

Kaunas, 2023

Kaunas University of Technology

Faculty of Social Sciences, Arts and Humanities

Human Performance in Machine Translation Post-Editing

Master's Final Degree Project

Translation and Localization of Technical Texts (6211NX031)

Sandra Vasiliauskienė

Project author

Prof. dr. Ramunė Kasperė
Supervisor

Assoc. prof. Jurgita Mikelionienė

Reviewer

Kaunas, 2023

Kaunas University of Technology

Faculty of Social Sciences, Arts and Humanities

Sandra Vasiliauskienė

Human Performance in Machine Translation Post-Editing

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Sandra Vasiliauskienė

Confirmed electronically

Vasiliauskienė, Sandra. Human Performance in Machine Translation Post-Editing. Master's Final Degree Project /supervisor prof. dr. Ramunė Kasperė; Faculty of Social Sciences, Arts and Humanities, Kaunas University of Technology.

Study field and area (study field group): N05 (Translation, Humanities).

Keywords: machine translation, post-editing, machine translation post-editing.

Kaunas, 2023. 48 p.

Summary

The topic of this master final degree project is the human performance in machine translation post-editing. The past decade has witnessed a notable evolution in machine translation (MT) technologies, which has highlighted the growing significance of MT and post-editing. Today, it is widely employed in both the industrial language business and the broader public. Advancements in technology coupled with the persistent demand for post-editing present translators with a noteworthy opportunity to enhance their skills and competencies. The assertion is made that translators ought to rely on technology to aid them in their professional duties, as it facilitates heightened levels of productivity and efficiency. The role of human translators has undergone significant transformation, as they are increasingly assuming the role of post-editors for machine translation. The performance, competencies, and abilities of post-editors are often a topic of discussion in both scientific and industrial contexts.

Diverse projects necessitate distinct specifications. The efficacy of a post-edited text may vary depending on the nature of the project it is intended for. The duration required for proficiently post-editing a text is subject to significant variation based on the project specifications, despite the consistency of the source material, unedited machine-translated text, and the human post-editor. The aim of the present investigation is to evaluate the efforts of post-editing in machine translation post-editing procedures. In order to achieve this goal, three specific objectives were established: to provide a comprehensive review of the existing literature on the quality of machine translation output and the performance of post-editing; analyse the user activity data with a focus on temporal and cognitive efforts; indentify the primary issues encountered during post-editing of two distinct types of content texts; compare to determine the post-editing strategies based on Blain et al’s typology (2011) utilized in post-editing procedures.

In order to accomplish the aim of the project, this research paper conducts a comparative analysis, both quantitative and qualitative of user activity data to identify the temporal and cognitive efforts involved, as well as the primary issues encountered during the post-editing of two distinct types of content texts. The study involved the utilization of the prevalent DeepL translation software to translate two distinct paragraphs extracted from the official Apple website. The evaluation of the machine translation output quality and the extent of post-editing modifications was conducted through the application of the Translog-II, an automated tool for assessing user data activity, and Blain et al's (2011) typology. Upon conducting a comparative analysis between the two texts, it was observed that the first text required an additional four minutes for post-editing, despite the fact that the Translog-II software recorded a lower frequency of events. The second text exhibited a mean execution time of 20 minutes, despite featuring a greater number of recorded events. In brief, it is conceivable that post-editors exhibit a greater degree of competence when translating instructional

manuals. Another factor that must be considered is that additional modifications were implemented due to inadequate results obtained from machine translation. The findings of the analysis indicate that the task of post-editing the user guide text was a complex endeavor, as evidenced by the implementation of a total of 188 modifications. In contrast to the primary text, a grand total of 121 alterations were implemented. The initial text underwent modifications primarily in the field of noun phrase selection, whereas the subsequent text exhibited a greater number of alterations pertaining to verb phrase selection. The findings of the study suggest that a significant proportion of the post-edited actions (PEA) recorded can be attributed to alterations made in the "Addition" strategy. In conclusion, it can be asserted that the task of post-editing an Apple website research article demanded a greater temporal effort, whereas the task of post-editing a user guide required a higher level of cognitive effort.

Vasiliauskienė, Sandra. Žmogiškieji veiksniai mašininio vertimo postredagavimo procese. Magistro studijų baigiamasis projektas / vadovė prof. dr. Ramunė Kasperė; Kauno technologijos universitetas, Socialinių, humanitarinių mokslų ir menų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): N05 (Vertimas, Humanitariniai mokslai)

Reikšminiai žodžiai: mašininis vertimas, postredagavimas, mašininio vertimo postredagavimas.

Kaunas, 2023. 48 p.

Santrauka

Šio magistro baigiamojo darbo tema – žmogiškieji veiksniai mašininio vertimo postredagavimo procese. Per pastarąjį dešimtmetį sparčiai vystėsi mašininio vertimo technologijos, o tai išryškino didėjančią mašininio vertimo ir postredagavimo svarbą. Šiandien pstredagavimas plačiai naudojamas tiek vertimo versle, tiek plačiojoje visuomenėje. Technologijų pažanga kartu su nuolatine postredagavimo paklausa suteikia vertėjams ypatingą galimybę tobulinti savo įgūdžius ir kompetencijas. Teigiama, kad vertėjai turėtų pasikliauti technologijomis, kurios padeda jiems atlikti profesines pareigas, nes jos padeda didinti produktyvumą ir efektyvumą. Vertėjų vaidmuo labai pasikeitė, nes jie vis dažniau atlieka mašininio vertimo redaktorių vaidmenį. Postredaguotojų veiklos rezultatai, kompetencija ir gebėjimai dažnai tampa diskusijų tema tiek moksliniame, tiek industriniame kontekste.

Skirtingi projektai reikalauja skirtingų specifikacijų. Postredaguojamo teksto veiksmingumas gali skirtis priklausomai nuo projekto, kuriam jis skirtas, pobūdžio. Teksto redagavimo trukmė gali labai skirtis priklausomai nuo projekto specifikacijų, nepaisant to, kad pirminė medžiaga, neredaguotas mašininio vertimo tekstas ir žmogus, redaguojantis tekstą, yra vienodi. Šio tyrimo tikslas - įvertinti mašininio vertimo postredagavimo eigos pastangas. Šiam tikslui pasiekti buvo iškelti trys uždaviniai: pateikti išsamią esamos literatūros apie mašininio vertimo rezultatų kokybės ir postredagavimo apžvalgą; išanalizuoti naudotojo veiklos duomenis, daugiausia dėmesio skiriant laiko ir kognityvinėms pastangoms; identifikuoti pagrindines problemas, su kuriomis susiduriama poredaguojant dviejų skirtingų tipų turinio tekstus; palyginti juos, siekiant nustatyti postredagavimo strategijas, remiantis Blain ir kt. tipologija (2011), naudojamas postredagavimo eigoje.

Siekiant įgyvendinti projekto tikslą, šiame moksliniame darbe atliekama kiekybinė ir kokybinė naudotojų veiklos duomenų lyginamoji analizė, siekiant nustatyti laiko ir kognityvines pastangas bei pirmines problemas, su kuriomis susiduriama postredaguojant dviejų skirtingų tipų turinio tekstus. Atliekant tyrimą buvo naudojama populiari DeepL vertimo programinė įranga, skirta dviem skirtingoms pastraipoms, paimtoms iš oficialios „Apple“ svetainės, versti. Mašininio vertimo kokybė ir postredagavimo atliktų pakeitimų apimtis buvo vertinama taikant Translog-II programą ir Blain ir kt. (2011) tipologiją. Atlikus abiejų tekstų lyginamąją analizę pastebėta, kad pirmajame tekste postredagavimas papildomai užtruko keturias minutes ilgiau, nepaisant to, kad Translog-II programinė įranga fiksavo mažesnį įvykių dažnį. Antrojo teksto vidutinė postredagavimo trukmė buvo 20 minučių, nepaisant to, kad jame užfiksuotas didesnis atliktų pakeitimų skaičius. Trumpai tariant, galima manyti, kad postredaktoriai, verčiantys naudotojo vadovą, pasižymėjo didesne kompetencija. Analizės išvados rodo, kad naudotojo vadovo teksto postredagavimo užduotis buvo sudėtinga - tai rodo iš viso 188 modifikacijų atlikimas. Priešingai nei pirminiame tekste, iš viso buvo atlikta 121 pakeitimas. Pirminiame tekste daugiausia buvo atlikta pakeitimų daiktavardinių frazių

parinkimo srityje, o vėlesniame tekste buvo daugiau pakeitimų, susijusių su veiksmažodinių frazių parinkimu. Tyrimo rezultatai rodo, kad didelę dalį užfiksuotų postredagavimo veiksmų galima priskirti pakeitimams, atliktiems taikant „Papildymo“ strategiją. Apibendrinant galima teigti, kad „Apple“ interneto svetainės mokslinio straipsnio postredagavimo užduotis pareikalavo didesnių laiko pastangų, o naudotojo vadovo postredagavimo užduotis pareikalavo didesnių kognityvinių pastangų.

Table of contents

List of figures

List of abbreviations and terms

Abbreviations:

MT – Machine translation

NMT – Neural machine translation

SMT – Statistical machine translation

TQ – Translation Quality

TQA – Translation quality assessement

PE – Post-editing

PEA – Post-editing actions

TM – Translation memory

MTPE – Machine translation post-editing

Introduction

The evolution of machine translation (MT) technologies over the past decade has underlined the increasing emphasis on MT and post-editing. Today, it is widely employed in both the industrial language business and the broader public. Improvements in technology and the ongoing need for post-editing provide translators with a significant chance to develop their skills and competencies. It is argued that translators should trust technology to assist them in their work, since it enables them to be more productive and efficient. Since human translators are increasingly becoming post-editors for machine translation, it is obvious that the function of the translator has changed dramatically. As a result, post-editor's performance, competences and abilities are frequently debated in the scientific and industrial areas. Despite the availability of high-quality MT technologies, the human touch will remain indispensable in translation.

The earliest attempts at machine translation emerged in the first half of the 20th century. In his account of the history of MT, Hutchins states that in 1933 the Russian Smirnov-Troyanskii devised the first important translation method. At the time, the initiative was substantial yet unknown outside of Russia. Hutchins notes that the first public display of an MT system occurred in January 1954 as a consequence of a partnership between Dostert and IBM, when 49 Russian phrases were translated into English utilizing a vocabulary of just 250 words and six grammatical rules (Hutchins, 1986). Despite having little scientific merit, this experiment inspired additional studies in other nations. This marked the beginning of the evolution of machine translation, which continues to this day. MT is not yet sophisticated enough to outperform human translators. Unquestionably, the rise of automated translation has made many human translators anxious they would be unemployed within a few years. Today, many translators share this sentiment. As MT quality is not good enough, there is often a need of post-editing, but there is no common standard of quality for post-editing (O'Brien, 2014).

Different projects require different specifications. The same post-edited text may be useless for one type of project while being suitable for another. The time necessary to post-edit a text successfully will change dramatically when project requirements are different, even if the source, raw MT text, and human post-editor are the same. The aim of this study is to assess machine translation post-editing efforts in post-editing processes. To accomplish this aim, four objectives were established:

1) to overview the literature on the machine translation output quality and post-editing performance;
2) to analyse the user activity data from the aspect of temporal and cognitive efforts;
3) to identify the primary problems encountered during post-editing of two different types of content texts;
4) to compare the results to determine the post-editing tactics according to Blain et al.'s (2011) typology used in post-editing processes.

There exists a limited number of experimental investigations in Lithuania utilizing Translog-II software, which have primarily employed it for the purpose of recording and analyzing eye-tracking outcomes. The current state of the research involves the collection of users' keylogging events, including insertion, deletion, copying, and other related actions. The study conducted an experimental analysis of two texts using a linear view, statistical data, and Blain et al.'s (2011) typology to obtain the findings.

1. Machine translation quality and post-editing performance

The rate of technological progress is incredibly great and rapid. Every day humanity faces a formidable challenge in the form of new technologies that alter the way people live and work. In the past, it was inconceivable that so many programs and tools could exist. It was intended to be a threat to someone's life and work. At present, they are an integral part of our lives and make our work easier. Over the last decade the emergence of machine translation technology has highlighted the new focus on MT and post-editing. Despite the high-quality MT tools, the human factor will always be vital in translation. Technology improvements and continuing demand for post-editing is an important opportunity for translators to improve their skills and competences.

1.1. Defining translation quality

Although translation quality assessement (TQA) is acknowledged as a major topic in the field of translation and localisation, the definition of translation quality (TQ) varies greatly across research and industry sectors. Since the second half of the 20th century, the concept of quality and the methods for determining it have been at the center of much discussion as the concep "quality" itself is simply too complicated and context-dependent. It has to deal with a variety of aspects, such as project specification and requirements, end-user's expectations, efficiency, source text complexity, availability of materials and etc. These aspects that may be assigned to quality cannot be given the same frame in every translation task, and are therefore differently visible or evaluable. Therefore, theoretically and practically, there appears to be no consensus on how to define quality. Furthermore, Horguelin and Brunette as cited in Martinez (2014) argue that there are numerous experts who continue to feel that translation quality is a relative and subjective idea (Martínez, 2014). Different scholars have diverse definitions of TQ. House (1996), for instance, argues that translation quality is evaluated in a variety of ways due to varying perspectives on it, and she outlines a number of approaches for evaluating TQ which are represented in the Figure 1.

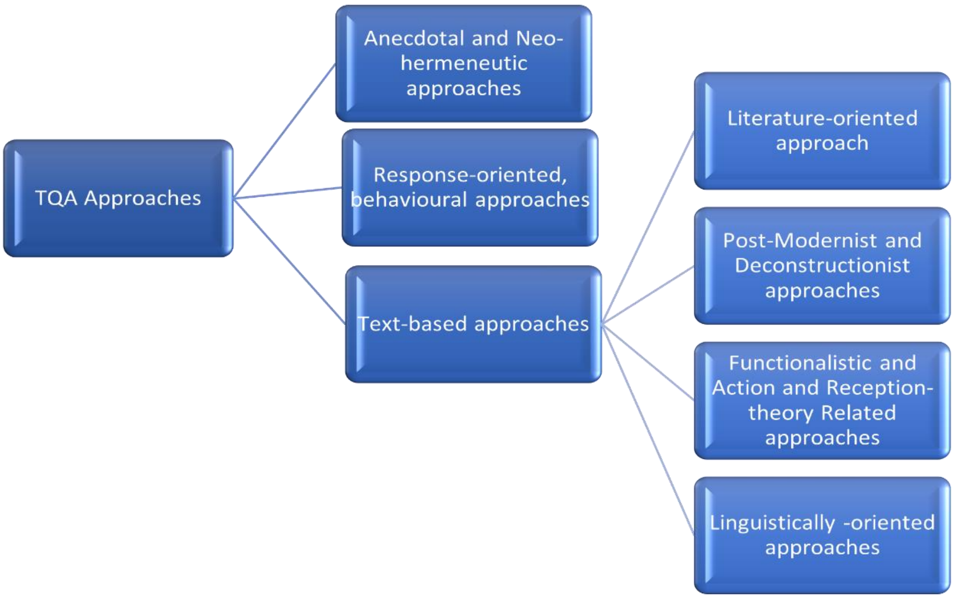

TQA Approaches
Anecdotal and Neo-hermeneutic approaches
Response-oriented, behavioural approaches
Text-based approaches
Literature-oriented approach
Post-Modernist and Deconstructionist approaches
Functionalistic and Action and Reception-theory Related approaches
Linguistically -oriented approaches

Fig. 1 Approaches for Translation Quality Assessment (House, 1996)

The anecdotal approach affirms that the quality of the translation depends on the translator's subjective interpretation, which is based on their knowledge and experience, whereas the neo-hermeneutic approach asserts that a good translation is the one where the translator identifies himself fully with the text he is translating. Behavioural or response-oriented approach is based on Nida's criteria: general efficiency, equivalence of response, and comprehension of intent. Text-based approach may be influenced by linguistics, sociology, philosophy, comparative literature, theories of action and reception (House, 1997). For some academics, the primary measure of a translation's quality is its accuracy, while for others, its fluency in the target language is more significant. Koby et al. (2014) state that no consensus has been reached on a definition of TQ and present two contrasting definitions as "broad" and "narrow". The broad and narrow definitions are presented as follows:

> "A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs."

> "A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture" (Koby et al., 2014).

The broad TQ perspective on translation necessitates that providers collaborate with the requester to comprehend their needs and expectations. The narrow TQ implies that translation assignments do not always require explicit specifications. The other academic Garvin (1984) presents his definition of quality which is based on the product. He outlines five major definitional approaches: transcendent, product, user, production and value. In addition, the scholar identifies eight quality dimensions (performance, features, reliability, conformance, durability, serviceability, aesthetics, and perceived quality) when a product can be rated highly on one quality dimension but poorly on another (Garvin David, 1984). There are likely many more attempts to describe TQ, yet despite these attempts, TQ definition remains elusive and it is evident that it is a matter of agreement between the researchers, requestors, providers, and end-users of the targeted project. The definition of quality within various academic disciplines is a topic of ongoing debate, as it can be assessed through a variety of objective and subjective views. So what is the definition of quality? The question of whether a translation task can be deemed qualitative or not is subject to debate. The outcome is contingent upon one's perspective and various intrinsic and extrinsic factors.

1.2 Machine translation output quality

The origins of computer-assisted translation (CAT) trace back even deeper than the 1980s. The initial systems were translation memories (TM) that saved human translations in a table and retrieved prior translations, either as an exact match or as a suggestion for a similar text. Beginning in the late 1940s, MT systems emerged along with TM. Initially, these systems were rule-based machine translation (RBMT) and required hand-coded linguistic rules. Later, statistical MT (SMT) or phrases-based MT evolution had a great impact on machine translation progress influencing the birth of neural MT (NMT). At present NMT, a form of statistical machine translation, is the state-of-the-art (Burchardt et al., 2021). NMT is now widely used in the translation industry because it outperforms statistically-based MT systems in a revolutionary way. It employs neural networks with the ability to learn complex linguistic patterns and provide more accurate and natural translations.The assessment process of machine translation output is a crucial yet challenging task. According to Popović (2011), this is because there is no definitive reference translation for a text, it is difficult to define automatic consuming. When determining which of translation systems seems to be the best, a great deal of effort has been devoted to locating metrics that correlate well with human evaluations (Popović & Ney, 2011). The development of computer-assisted translations has undergone a lengthy process since the 1980s and continues to progress alongside advancements in improving translation memory, artificial intelligence, and ongoing academic studies and experiments.

The output of machine translation can be examined manually, automatically, or through a combination of both methods. A computer uses an algorithm that is encoded in a software to determine the evaluation score during the automatic evaluation. The determined score informs the user of the quality of the translation (Kasperavičienė et al., 2020). In addition, the output of MT might be analyzed from several angles and perspectives. Applying automatic quality estimation metrics, conducting an error analysis by professionals, employing cognitive scientific method with human experts, semiexperts or non-experts, determining the acceptability of the output of non-experts/non-professionals/amateur users via qualitative methods are all ways to evaluate the overall quality of machine translation (Kasperė et al., 2023).

There are a variety of methods and approaches for evaluating translation quality, therefore a survey of the scholars' efforts in this area might be illuminating. As cited in Dabbadie et al. (2002) Halliday created a four-point intelligibility scale, ranging from 0 to 3, with 3 representing the most comprehensible. Vanni and Miller consolidated the International Standards for Language Engineering (ISLE) project's suggested metrics of comprehensibility, readability, style, and clarity into a single assessment factor: clarity. Hutchins and Somers (1994) and Arnold et al. (1994) claimed that intelligibility (or clarity) is a reasonable measure of translation quality; however, only individual phrases are assessed, making it more subjective and unclear than if entire texts were graded. It is obvious that machine output quality varies based on the approaches, time, and content used (Dabbadie et al., 2002).

Other academics Maučec and Donaj (2019) state that traditionally, there are two approaches for evaluating machine translation: glass-box evaluation and black-box evaluation. Glass-box evaluation analyzes a system's quality based on its internal attributes. Black-box evaluation assesses simply the system output without considering inner processes of the translation system. Automated evaluation criteria, such as BLEU, NIST, and WNMf have been created to quantify the similarity between the output of MT and professional human translations. However, the vast majority of research on MT assessment focuses on analyzing raw MT output as opposed to post-edited text, and it is argued that comparing raw MT output to the final form of a human translation is an unfair comparison (Maučec & Donaj, 2019). According to Hutchins and Somers, the most evident measures of translation quality are: fidelity, intelligibility and style. Intelligibility and fidelity are the human evaluation measures of MT output quality that are most often employed (Fiederer & O'Brien, 2009). Evidently, the most crucial aspect is that there are numerous ways and strategies for measuring translation quality, and it is quite hard to figure out which method or approach is the best and most accurate.

1.2.1 Manual (human) evaluation

Human review of contemporary machine translation systems is a challenging subject. Although many studies have been conducted on human evaluation , the subject lacks a universally approved standard technique. This might be mainly because manual evaluation relies on the subjective analysis and evaluation of an expert, which is influenced by many circumstances. The other issue is what and how is going to be evaluated. Human evaluation methods include direct assessment, ranking, task-based evaluation, and post-editing efforts. It is reasonable to presume that particular barriers and benefits, along with drawbacks, are associated with each form of human evaluation. Unquestionably, the aspect of agreement on what is " good " presents a number of difficulties.

The earliest human evaluation methods for MT date back to 1960s and include the Automatic Language Processing Advisory Committee's intelligibility and fidelity measures (ALPAC). In the 1990s, the Advanced Research Projects Agency (ARPA) developed a method for evaluating machine translation systems based on the adequacy, fluency and comprehension of the MT output. The evaluation of fluency is done similarly to the adequacy evaluation, with the exception that the evaluator must provide intuitive judgments on a sentence-by-sentence base for each translation. The results are calculated by averaging the evaluations over all translation set decisions. White and Taylor (1998) created a task-oriented evaluation technique for Japanese-to-English translation in order to evaluate MT systems based on the tasks for which their output could be employed (Han, L., Jones, & Smeaton, 2021). According to King et al. (2003) in accordance with ISO 9126 standards (ISO/IEC 1991), the sole quality characteristic considered pertains to a singular subcharacteristic of functionality, namely accuracy. This subcharacteristic is defined as the ability to provide the correct or mutually agreed upon outcomes. In addition to accuracy, scholars extended the manual evaluation methods for MT systems adding suitability and interoperability to the other top-tier ISO qualities such as reliability, usability, efficiency, maintainability and portability (King et al., 2003). Voss and Tate (2006) presented task-based machine translation (MT) output evaluation by extracting who, when, and how.

Advancing further, Fomicheva quoting Linguistic Data Consortium (2005) asserts that manual MT is evaluated on a multi-point scale for translation adequacy and fluency. Adequacy assesses how much meaning of the source sentence is retained in the MT output. Fluency refers to the linguistic integrity of a translation. She claims that rather than the original text, human translation is frequently utilized to evaluate sufficiency. Because it does not require bilingual speakers, such a monolingual reference-based examination is a feasible and appealing option. Furthermore, she reports that some early techniques utilized evaluation frameworks designed for assessing the quality of human translation. Her study reveals that in addition to the inherent bias provided by the individual tastes and expectations of the annotators, it is proved that the reference translation has a systematic effect on monolingual rating. When a distinct human translation is utilized as the gold standard, annotators consistently assign different scores to identical MT results (Fomicheva, 2017). Specia et al. (2011) conducted a study on the adequacy of MT and classified it into four categories: highly adequate, fairly adequate, poorly adequate, and completely inadequate. The academics proposed the methodology that relied on human evaluations of adequacy and various translation quality metrics to compare the source and translated texts. The information varies from basic frequency data regarding tokens in both the source and target sentences to varying degrees of linguistic information (Specia et al., 2011) Graham et al. (2013) offered continuous measurement scales (CMS) employing fluency criteria. (Graham et al., 2013)

The development of new technologies encourages scientists to look for novel approaches and techniques. As a result, they continuously push the limits of what is feasible and explore new fields within their specialties. Popovic (2020) proposed a new technology that requires human assessors to mark all troublesome sections of candidate translations, whether they be words, phrases, or sentence. In her study annotators were instructed to discern only between big and minor concerns. Since it does not incorporate fine-grained mistake categorisation, inter-annotator agreement is strong; annotators concur on 70-80% of word-level issues. Some annotators tended to mark entire phrases/sentences, whereas others preferred to select only one or two words. On texts annotated in the manner outlined, it is also feasible to do a detailed study of disagreements, which could be an intriguing avenue for gaining a deeper knowledge of the evaluation process. The evaluation of translated English customer reviews into Croatian and Serbian was determined by two quality requirements: comprehensibility and adequacy (Popović, 2020). In a study of Freitag et al. (2021) they present an evaluation technique based on explicit error analysis and the Multidimensional Quality Metrics (MQM) model (Lommel et al., 2014). The study demonstrates that human assessments of crowdworkers (as performed by WMT) have a low correlation with MQM scores, leading to significantly divergent system-level ranks. This discovery raises doubts on prior conclusions based on human evaluation by crowdworkers, particularly for high-quality MT. Furthermore, the researchers declare that many automated measures, particularly embedding-based ones, currently exceed human evaluation by crowdworkers. In contrast to ratings gained by crowd-workers and ratings acquired by professional translators employing simpler human assessment approaches, MQM labels collected by professional translators reveal a significant quality disparity between human and machine-generated translations. This illustrates that human translations produced by professionals continue to outperform computer translations (Freitag et al., 2021). In a research by Han et al. (2021), human evaluation methods are divided into two categories: traditional and advanced. The first includes intelligibility, fidelity, fluency, adequacy, and comprehension, while the second includes task-oriented, extended criteria, post-editing, segment ranking, crowd source intelligence (direct assessment), and revisiting traditional criteria (Han, L. et al., 2021). Undoubtedly, there exist numerous alternative techniques, criteria and approaches for assessing human performance. In summary, it might be claimed that a variety of human evaluation techniques and strategies are employed in the development of artificial intelligence technology.

1.2.2.Automatic Evaluation

Manual evaluation has several disadvantages, such as being time-consuming, expensive and not duplicatable. Because of this, automatic metrics are widely utilized throughout the world. In most cases, these measures contrast the results of MT systems to human reference translations, yet there are also metrics that do not utilize reference translations. Automatic evaluation is generally acknowledged to be objective and cost-effective.

Since the earliest days of MT, the need to evaluate MT systems based on their output has existed, so it comes as naturally that MTE has been exhaustively investigated. In a study by Novikova et al. (2017), automatic metrics are divided into word-based metrics (WBMs), e.g. BLEU (Papineni et al., 2002), TER (Snover et al., 2006), NIST (Doddington, 2002), LEPOR (Han et al., 2012), METEOR (Lavie & Agarwal, 2007) and ROUGE (Lin, 2004) and grammar-based metrics (GBMs). A comprehensive research examination of errors indicated that automated metrics exhibited limited efficacy in discriminating between outputs of moderate and high quality, possibly due to the incongruity between the scales of human evaluations and metrics. The study reveals that word-based metrics rely on two human-generated references and are regarded as a correct and comprehensive gold standard. In contrast to reference-based metrics, grammar-based metrics are independent of human-generated references and remain unaffected by their quality (Novikova et al., 2017).

Other researchers utilize a diverse classification system. For instance Kocmi et al. (2021), rely on two distinct types of metrics for automatic machine translation: string-based metrics and metrics utilizing pretrained models. BLEU, EED (Stanchev et al., 2019) , ChrF (Popović, 2015), CharacTER (Wang et al., 2016) and TER are single-based measures, whereas BERTScore (Zhang et al., 2019), COMET (Rei et al., 2020), BLEURT (Sellam et al., 2020) are pretrained metrics. The former compares the coverage of diverse substrings between the human reference text and the MT output text. The final category of pretrained approaches includes metrics that employ neural models that have been pretrained to evaluate the quality of MT output texts given the source sentence, the human reference, or both. They are not strictly rely on the translation quality of the human reference. Furthermore, the researchers share evidence indicating COMET being the most effective performance metric, whereas ChrF is the most effective string-based algorithm (Kocmi et al., 2021).

Han et al. (2021) classify metrics into traditional and advanced with the following categories: n-gram word surface matching, deeper linguistic features and deep learning models. According to the rsearchers, the straightforward n-gram word surface matching techniques emphasize accurate surface word matches in the output translation. The metrics that fall within the first group (simple n-gram word matching) have the benefit of being effective at quantifying translation fluency, are very fast to compute and have low cost. However, the syntactic information is hardly taken into account (Han, L. et al., 2021).

According to Chauhan and Daniel (2022), automatic certain MTE measurements share traits with other categories, making it difficult to classify them. Edit distance, precision, recall, F-measure, and word order are lexical similarity techniques. The character-based evaluation measures include precision and recall for character n-grams. Parts of speech syntactic qualities include speech tags, phrase types, and sentence structures, whereas named entities, synonyms paraphrasing, semantic, and language models are among the semantic features. In addition, the researchers provide a comprehensive study presenting properties, advantages and disadvantages of each metrics (Chauhan & Daniel, 2022).

Marie’s (2022) report offers an automated evaluation of the comprehensive machine translation endeavor of the Seventh Conference on Machine Translation (WMT-22). A thorough assessment was carried out on 185 systems, covering 21 translation directions. The evaluation included translations between languages with abundant resources and those with limited resources, as well as translations between languages that are closely related and those that are distant. The current research's comprehensive automated evaluation highlights specific limitations of modern machine translation systems. The aforementioned text showcases the capacity of automated metrics, namely chrF, BLEU, and COMET, to supplement one another in order to overcome their individual shortcomings with regards to precision and comprehensibility (Marie, 2022).

The aforementioned data indicates that the assessment of various metrics is a topic of debate. Certain scholars assert that COMET and ChrF metrics are superior, whereas others contend that their scope is restricted. The primary reason for this disparity is the varying methodologies and techniques employed by distinct research studies in assessing metrics.

1.3. Machine translation post-editing performance

MT is rapidly becoming faster, more affordable, and more accurate. Yet, it has a long way to go before it is on level with human translators. Machine translation post-editing (MTPE) involves combining the best of each world: the speed and ability of MT engines to efficiently process texts, and the expertise and awareness of experienced linguists. MTPE is now the norm for businesses who wish to distribute their material in several languages to a worldwide audience. The process of selling a product necessitates its localization to the specific country or region it is intended for. This underscores the importance of comprehending MTPE in the localization process, irrespective of one's role in the translation industry.

MTPE is understood as “edit and correct machine translation output” (ISO 18587:, 2017), TAUS (2010) claims “post-editing being the process of improving a machine-generated translation with a minimum of manual labor” (Zaretskaya, 2017). According to Krings (2001), post-editing is typically defined as the human process of comparing a source text with the raw output of MT system and then modifying the output language text to match some criteria of acceptability for a particular purpose (Krings, 2001). O’Brien (2011: 197) defines post-editing as “the correction of raw Postmachine-translated output by a human translator according to specific guidelines and quality criteria“ (O’Brien, 2011). Based on the aforementioned definitions of post-editing, it can be inferred that post-editing (PE) refers to the process of revising machine-generated translations by a human translator.

1.3.1. Theoretical background: then and now

Machine translation dates back to the invention of the digital technologies, although MTPE is a relatively new innovation. In reality, less than a decade has flown by since MTPE was regarded a feasible alternative for industrial application. Garcia reports that in the late 1950s and the beginning of the 1960s, administrators and academics with a perspective of what language and technology could accomplish combined made postediting a popular topic. In the same year as the Georgetown experiment marked the beginning of large-scale funding in the United States, the journal Mechanical Translation was created. The first issue was devoted to assembling abstracts of previously published literature on MT, beginning with the 1949 Warren Weaver memorandum and ending with a thorough discussion of the 1952 First Conference on Mechanical Translation. In pre-MT literature, the concept of Mechanical Translation (MT) was considered with the goal of creating a readable translation in a fraction of the time it would take a bilingual expert. The initial instance of practical application was recorded at the RAND Corporation, where the individual responsible for editing the text was required to possess expertise in both the English language's grammatical rules and the subject matter of the article. Knowledge of the Russian language was only necessary for the linguists involved in the text preparation process before inputting the text into the machine. Between 1954 and 1966, the US government invested nearly $20 million in MT, providing the majority of the push for its development (Garcia, 2012).

According to Garcia (2012), the first postediting program was terminated in the early 1960s. The US Air Force's Foreign Technology Division and Euratom used postediting, but funding in the United States ceased in part due to a 1966 report by the Automatic Language Processing Committee (the so-called ALPAC report) that concluded postediting was not worth the time, quality, and difficulty compared to human translation (Koponen, 2016). The most significant aspects of MT and postediting are that they were originally utilized by the ALPAC in the 1970s and continued to be used in the United States, Canada, France, the Soviet Union, and Japan until the 1980s. By the middle of the 1980s, postediting would be performed on-screen, with the original and the translation displayed side-by-side, and posteditors would be required to acquire new abilities. The 1990s were the era of email and the World Wide Web, with US corporations employing Systran. Nitzke and Hansen-Schirra (2021) report that 1968s were the beginnings of Systran when IBM developed the first commercial statistical machine translation system (SMT) Systran which was established by Peter Tome and now it is supposed to be the one of the oldest MT companies (Nitzke & Hansen-Schirra, 2021). Heading forward to the development of MTPE, the European Commission utilized postal services for a total of 180,000 pages in 1995 from 30,000 pages in 1990.

The end of the 20th century witnessed a shift from postediting to MT. This new technology substituted PC-based machine translation and expanded the usage of MT, which is now accessible to the typical web user as opposed to just defense analysts and manual readers. Garcia (2012: 300) states that Miller and Beeve-Center introduced three evaluation methods in 1956: subjective scaling, comparing the test translation with a translation of granted excellence, and asking people who had only read the machine-translated version to answer comprehension questions. Orr and Small (1967) were the originators of postediting research, using multiple-choice questions to evaluate the readability and comprehension of Russian material translated by traditional ways and after machine output was post-edited. Yngve was the first one to give a comprehensive profile of a posteditor (Garcia, 2012).

By the mid-1950s, the primary characteristics of the posteditor had been established, and even though there was dispute regarding the type of individual a posteditor should be. In the sudy of Vieira (2019), it is claimed that the post-editing of MT has been a training, a service, and a focus of research from the earliest days of MT technology, but its significance has increased in recent years. It was viewed as a passive activity in which human editors helped reduce the gap between flawed MT outputs and the ultimate goal of MT, which is to provide entirely automatic, high-quality translations. Post-editors were considered to as MT's “human partners “ and did not need need to be fluent in the source language (Vieira, 2019). The 1970s were particularly barren for research, as MT and posting delayed the necessary technology. In the 1980s, research on postediting restarted alongside the introduction of MT by a number of significant corporations, the European Commission, and the Pan-American Health Organization. In the research by O'Brien and Simard (2014) it is mentioned that Krings’ work in the 1990s paved the way for experimental studies of post-editing, which were aided by new tools such as screen recording, keyboard logging, and eye tracking (O'Brien & Simard, 2014).

Nowadays recent advancements in machine translation compel translators to migrate from traditional translation to PE of machine-translated text, a technique that saves time and increases quality. Although MT has undergone great progress in recent years, an increasing number of professional translators incorporate the technology into their translation workflows. Herbig et al. (2019) declare that neural machine translation (NMT) paralleled and finally surpassed the success and popularity of statistical machine translation (SMT) systems. As quality has improved, the usage of MT and PE in professional translation processes has also expanded. The majority of professional translators utilize CAT (computer-assisted translation) tools today. The academics state these include MT and TM along with quality estimation and concordance functionality alignments between source and MT color coding to indicate the correlation between input sentences and TM matches. Furthermore, the authors indicate the future research on translation environments should focus more on mouse and keyboard-based techniques. In contrast, eye tracking and gesture modalities appear less significant (Herbig et al., 2019).

Despite the fact that MT has advanced to the point where it can now fully assist the work of a human translator, there are still situations in which it cannot be relied upon. Texts used for advertising or marketing purposes, for instance, are typically written with the objective of being amusing or hilarious and are intended for a large audience. The utilization of Machine Translation (MT) may result in the loss of subtleties such as humor, cultural references, metaphors, and other nuances due to the absence of the human's lifelong experience and social background. Similarly, this assertion can be applied to texts that require specialized knowledge in a particular field, such as medicine, law, or engineering. In this case, post-editing constitutes an essential element of the translation process.

1.3.2. Guidelines and strategies for post-editing

The current investigation focuses not only on the quality of machine translation but also on the analysis of post-editing efforts. Therefore, a comprehensive examination of sources is imperative to gain familiarity with all guideline recommendations and strategies. Multiple academics, IT specialists, and developers of CAT tools provide post-editing success-enhancing PE strategies for rendering post-editing efficient. It shoud be mentioned these are just guidelines but not standard. In the research of Hu and Cadwell (2016) the comparison of five proposals (O’Brien, 2010; Mesa-Lao, 2013; Flanagan and Christensen, 2014; Densmer, 2014; TAUS, 2016) is analysed. The comparative analysis reveals that the existing PE rules contain numerous overlaps, particularly for light post-editing. The primary variations are found in the full PE rules and address the necessity for style and the expected quality of the target content, which varies depending on the intended use of the material (Hu and Cadwell, 2016).

TAUS established in 2005 as a translation automation and innovation polling company which began by advocating the straightforward notion that machine translation is a helpful tool for the translation industry and underlined the need for innovation, open platforms, and cross-industry cooperation. The rules of TAUS PE guidelines, which were developed in 2010, begin with two basic criteria and advice for minimizing the amount of post-editing required. In 2016, TAUS revised their PE guidelines to contain five sections, which are subdivided into four sections: evaluation post-editor performance, production, pricing machine disobedience PE guidelines, and about the MT guidelines. TAUS suggests an agreed set of recommendations in order to minimize the amount of post-editing needed:

- It is essential to fine-tune the system correctly, ensuring high-level dictionary and linguistic coding for rule-based machine translation systems or training with clean, high-quality, domain-specific data for data-driven or hybrid systems.
- It is important to verify that the original text is well-written (i.e., with correct spelling, punctuation, and ambiguity) and, if possible, optimized for machine translation.

- Integration of terminology management across the authoring, machine translation (MT), and translation memory (TM) systems.
- Education of post-editors beforehand.
- It is important to examine the quality of the raw MT product and establish appropriate expectations.
- Establishing a specification for the ultimate quality of the post-edited material based on user type and acceptance levels.
- Paying post-editors to provide structured comments on typical MT problems so that the system may be enhanced over time (Massardo et al., 2016).

Post-editing encompasses multiple stages, which entail verifying precision, rectifying grammatical and syntactical inaccuracies, and modifying the text to guarantee its lucidity and comprehensibility. The individual responsible for editing may also require the assistance of reference materials or consultation with experts in the relevant field to guarantee the precision and suitability of the translated text for its designated readership.

O'Brien presented a tutorial on post-editing at the 2010 AMTA conference and introduced the general PE guidelines of Wagner (1985) as well as the light and full post-editing recommendations. In accordance with Evolution Theory (AYA) and the Technology for Non-Tolerance, she enumerated the criteria that determine post-editing levels, such as the time of translation, life expectancy, and perishability of the material. She noted that light post-editing was not a simple task for linguists due to the fact that they had to ignore probable   which refer to the necessity for full post-editing at the cost and speed of light postedting (Hu & Cadwell, 2016). Furthermore, numerous corporations propose various suggestions. One proposes primary training objectives and post-editing recommendations, which are itemized as follows:

- Increasing awareness of MT in the linguistic community as a whole. MT is seen as an additional productivity tool as opposed to a technology designed to replace human linguists. Human skill is essential for producing a translation of high quality.
- Handling the linguists’ expectations regarding the quality of machine translation and the assignment itself. Several linguists anticipate that MTPE will be comparable to human translation proofreading. Yet, both the quantity and character of MT errors differ from those found in human translations. So, the work and time required to post-edit MT documents differs from that required to review a text supplied by a human translator.
- Describing what is expected of the linguists in terms of the quality of the final translation. Like with many other LSPs, we distinguish between Light and Full post-editing for MTPE quality.
- Giving advice and suggestions on how to improve post-editing efficiency. There are a lot of time-saving post-editing strategies that can be utilized.
- Demonstrating to the linguists how to provide constructive criticism to improve the quality of MT (Zaretskaya, 2017).

Consequently, we can observe that the task of post-editing is gaining significance within the translation industry due to the growing adoption of machine translation technology by various organizations. This technology is being used to enhance translation efficiency and minimize expenses. Postediting can be a laborious task that demands specialized expertise and knowledge.

Therefore, it is crucial to collaborate with proficient posteditors who possess the necessary training and proficiency in the most effective techniques for postediting machine-translated content.

According to O’Brien et al. (2014), in the study of Rico and Ariano (2014) PE guidelines are divided into two sets: language independent (LI) and language specific. The set of language independent (LI) guidelines refer to the following:

- Fix any wrong term in the text, either technical or non-technical. Correct also any inconsistent use of the same term.
- Fix any syntactic error which consists of wrong part of speech, incorrect phrase structure, wrong linear order of words and phrases.
- Fix any morphological error which consists of wrong morphological form (number, gender, case, person, tense, mood, voice, aspect).
- Fix any missing text (paragraph, sentence, phrase, word) as long as the omission interferes with the message being transferred.
- Fix any misspelling.
- Fix incorrect punctuation as long as it interferes with the message.
- Do not fix stylistic problems, unless they interfere with the message.
- Fix any offensive, inappropriate or culturally unacceptable information (O'Brien et al., 2014).

Despite the lack of standard for PE, the taxonomy of post-editing strategies is evident with the exception of a few recent scholarly proposals. The preceding literature demonstrates that the recommended machine translation post-editing strategies for professional translators generally adhere to the framework of binary division, consisting of light MTPE and full MTPE. Light post-editing and full post-editing are well-known and extensively used techniques. TAUS (2016) presents two post-editing quality levels: “good enough” and “publishable quality”. The first level of quality is characterized as comprehensible and accurate, conveying the sense of the source text without being necessary grammatically or stylistically flawless. At this stage, the post-editor must ensure that the translation is semantically accurate, that no information was added or missing by accident, and that it does not contain any objectionable or inappropriate material. The second level would be comparable to human translation expectations. In addition to being comprehensible and correct, the writing should be correct grammatically and professionally. This classification takes into account the editor's efforts and the degree of editing, but disregards the textual functions of various MT text types (Massardo Isabella et al., 2016)

Chung-ling Shih (2021) suggests dividing strategies into three categories: “accurate-enough editing”, “clear-enough editing” and “attractive-enough editing” taking into consideration type of the text. The academic believes that while modern NMT methods have vastly increased the accuracy of translation, only linguistic MTPE of the most fundamental level is required for the publishing of technical texts. Yet, journalistic and business web content necessitate pragmatic MTPE for coherent, communicative translations. Affective MTPE facilitates the creation of an emotional appeal for the purpose of marketing, although its use is optional. For product instructions and user manuals that aim to convey accurate information, post-edited MT texts must be semantically and grammatically accurate; for journalistic texts that aim to present clear, comprehensible content, post-edited NMT texts must be as clear and communicative as possible; and for web-based company texts that aim to inform and attract audiences, post-edited NMT texts must be as communicative and appellative as possible (Chung-ling Shih, 2021).

Similar to the preceding classification, we may observe it in Allen's (2003) research. The scholar classifies post-editing into three levels as well: rapid PE, partial PE (or minimal PE) and full PE. It could be said rapid PE (RPE) mainly was created for European Commission (EC) as there was need for quick translation adjustments of MT output, for urgent materials meant for informational reasons or restricted distribution, such as working papers for internal meetings, meeting minutes, technical reports or annexes, etc. In general, the purpose of RPE is to conduct a rigorous minimum of repairs on documents that often include perishable information (Allen, 2003). In industry and business in the 1990s, the phrase "minimal PE" became common (Allen, 2003).

Despite numerous guidelines, recommendations, and post-editing procedures, post-editors frequently encounter certain issues and problems. According to TAUS (2016), light post-editing problem areas are: (1) properly expressing the meaning of the original text; (2) fixing terminology inconsistencies; (3) eliminating duplicates and rectifying omissions (for SMT output post-editing); (4) negations, word order, singular vs. plural and morphology. The examples of recognized full post-editing issue areas include: (1) handling of measures and locale-specific punctuation, date formats; (2) resolving terminology inconsistencies and terminology disambiguation; (3) handling list elements, tables, or headers as opposed to body content; (5) handling proper names, product names, and other not translatable components;(6) repetitions (constant precise matches); (7) eliminating duplicates and correcting omissions (for SMT output post-editing); (8) morphology (agreement), negations, word order, and singular vs plural (TAUS, 2016). Despite varying recommendations and strategies, it can be asserted that post-editing is receiving increased attention within the discipline of translation studies. This is due to a growing recognition of its significance for maintaining the quality of machine translation output.

1.3.3. Post-editing functionality: temporal, cognitive and technical efforts

Despite the fact that MTPE has been a topic of practice, service, and study for many years, it is continuously evolving and undergoing development and therefore reseachers focus on different PE aspects: machine translation raw output, translation quality, post-editor's performance, PE efforts and etc. The interdependent relationship among effort, speed, and quality in translation necessitates the need for researchers to identify and develop methodologies for comprehending and quantifying the effort expended by translators in their professional capacity.

Krings (2001) identifies three aspects of post-editing: temporal, which refers to the time spent on post-editing, cognitive, which refers to detecting problems and the necessary actions to remedy them, and technical, which refers to the edit operations conducted to generate the post-edited version. In different settings, these aspects of effort are not always equivalent (Krings, 2001). Post-editing effort is typically measured in terms of post-editing time due to the importance of time in the process. In the fast-paced translation industry, time is a crucial component of the post-editing process. Later on, O'Brien (2007) develops a research based on the Krings' (2001) that analyzes the temporal and technical effort in the post-editing process. In her work, she used the notion of NTIs (Negative Translatability Indicators), which refers to "linguistic features known to be problematic for MT ". In this study, O'Brien examines temporal and technical effort in portions of the source text that contain NTIs and compares them to parts from which these indicators have been deleted (O'Brien, 2007).

In the research of Dede (2022), the interesting findings are presented. The results showed that the participants spent more time on human translations than on machine translations, which is an intriguing conclusion. The human translations used in the experiment were extracted from a corpus of reference translations and modified to seem like fuzzy matches (Dede, 2022). Even though time might be thought of as the most obvious component of post-editing effort, Krings (2001) claims that it is not always simple to get precise data on post-editing time in real-world work environments. For instance, post-editors' self-reports of the time spent may not be thorough or adequate, and accurately capturing information would require specialized technologies that are not always available (Koponen, 2016).

Cognitive effort refers to the mental exertion required to read the materials, consider how to translate, and fix errors in translation. A post-editor must expend greater cognitive effort when confronted with more challenging content that has been less successfully machine translated (Lacruz, 2017). Generally, it looks difficult to quantify cognitive effort. Krings (2001) employed think-aloud protocols (TAP), in which post-editors reported their activities verbally during the post-editing process. This technique has been frequently utilized in the past to study the translation process, but it has numerous downsides, such as slowing down the process and altering the cognitive processing involved. Using TAP, pause measurement, and, increasingly, eye-tracking, cognitive effort has been quantified (Krings, 2001).

Due to prominent research such as O'Brien's (2006) pilot study on fuzzy match editing and post-editing work, eye-tracking has become a popular technique for evaluating cognitive effort in translation studies. Furthermore, eye-tracking has been utilized in a number of studies to quantify cognitive work during post-editing. O'Brien (2011) required seven participants to post-edit 60 segments of English-French SMT output, 20 segments in each of three GTM (General Text Matcher, Turian et al., 2003) score categories, using the Alchemy Catalyst editing environment. She discovered that average fixation time per word and average fixation count per word were substantially connected with the GTM categories, indicating that the GTM metric may be an effective predictor of cognitive PE effort (O’Brien, 2011). Koglin (2015) had 14 translation students post-edit two writings regarding the Tea Party movement in the United States that had been translated from English to Portuguese using both Systran and Google Translate MT systems within the Translog-II environment. He discovered that post-editing required less cognitive work than translating the texts from scratch (Walker & Federici, 2018).

Moorkens et al. (2015) examined whether human estimations of PE effort were accurate predictors of actual PE effort and if post-editor behavior differed when PE effort estimation indications (based on actual user ratings) were shown to participants. There was a moderate correlation between measurements of PE effort and mean user ratings (six participants rated the segments that had been machine translated from English to Portuguese), leading to the conclusion that “human ratings of PE effort do not correlate strongly with the actual time required during post-editing “. The moderate correlation indicated that, as participants progressed through the texts to be post-edited, there was some relationship between the three-category, 'traffic light' indicator color scheme and the final measurements of temporal and technical effort (Moorkens et al., 2015).

Other researchers suchF as Alves et al. (2016), utilized the Casmacat interface to conduct an A/B test, inviting participants to post-edit with and without interactive machine translation (IMT) capabilities, in order to examine the effect of IMT on PE behavior. The MT recommendation is updated in real time depending on the user's modifications while IMT is engaged. The researchers speculated that technical and temporal effort would be reduced in the interactive PE mode, but did not make any cognitive effort estimates. In actuality, neither technical nor temporal effort decreased as anticipated, but the mean length of fixation was shorter than with standard PE (Alves et al., 2016).

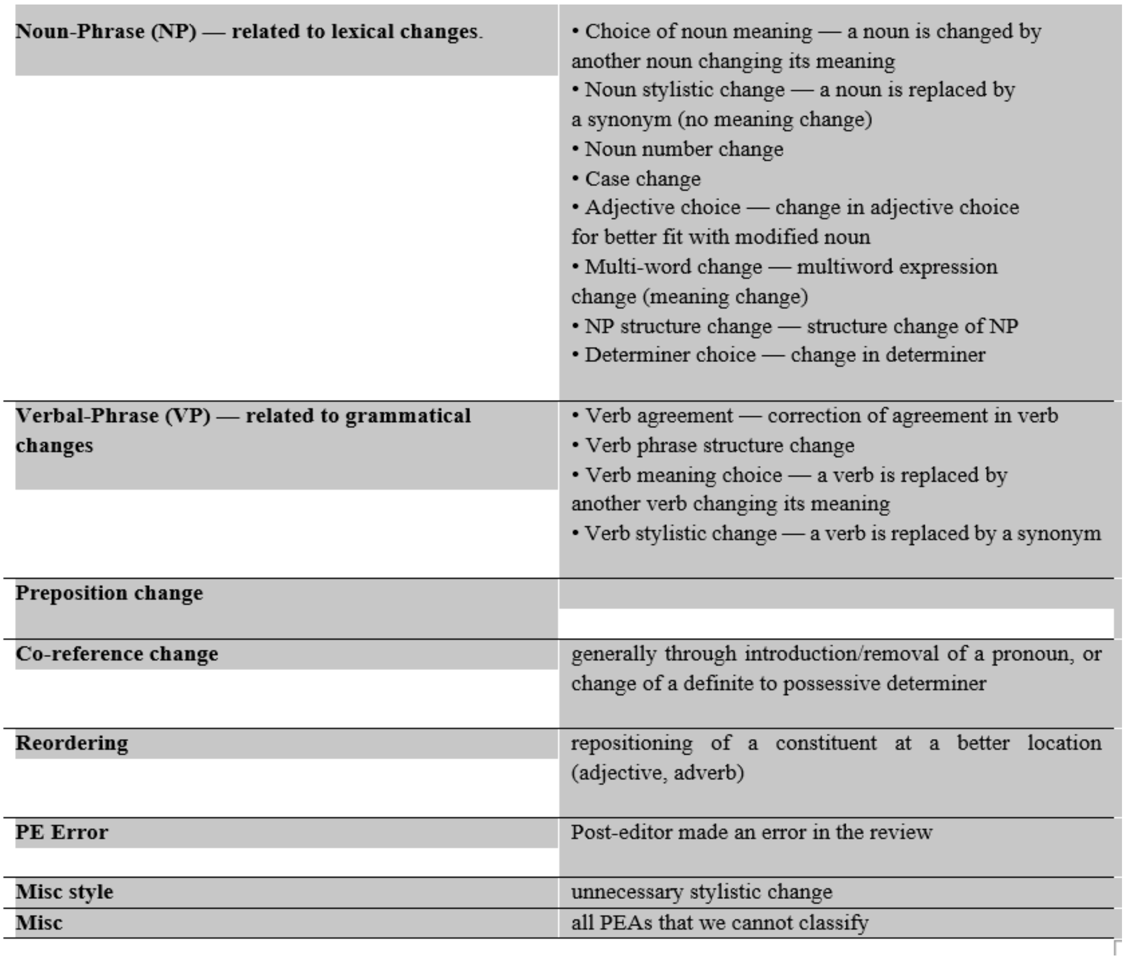
Based on the amount of keystrokes, mouse clicks, and eye fixations in a section, Laubli and Germann (2016) developed a statistical model for annotating PE. Ten experienced annotators were more accurate than the statistical model in comparison to a gold standard sample annotation of seven PE sessions, while two were less accurate. This is a promising outcome for automatic annotation, but it shows that data processing for eyetracking TPR data will remain a labor-intensive endeavor for the foreseeable future (Läubli and Germann, 2016).

The research of Colman et al. (2021) outlines the experimental design of an eye-tracking in which participants alternated between a machine translation (MT) and a human translation while reading the whole novel (Agatha Christie's The Mysterious Affair at Styles) in Dutch with the aim to examine the reading processes of individuals reading both versions, determine the extent to which MT influences the reading process and to find out which faults have the most influence on this reading process (Colman et al., 2021).

In the study of Kasperė et al. (2023) the experiment included an eye-tracking, a questionnaire, and a large-scale population survey to assess the goals, typical conditions, and systems utilized by non-professional users of machine translation. Eye tracking was conducted utilizing a commercial non-invasive eye tracking equipment and SMI BeGaze 3.7.2.42 software for data processing. The results of the study indicate that the average fixation time was longer on areas of interest with errors than on areas of interest without errors, confirming the findings of other studies that errors attract more attention of the readers and require more cognitive effort than correct text. The average fixation time on regions of interest with mistakes (as a percentage of overall trial time) was 12.6% for expert users of machine translation and 11.7% for non-professional users. On the contrary hand, professionals exhibited a longer average fixation period on error-free regions of interest than non-professionals, i.e. 11.8% versus 10.4% (Kasperė et al., 2023).

Technical effort is the amount of real edits conducted by the post-editor, which may be estimated using the HTER metric created by Snover et al. (2006), which determines the fewest feasible changes necessary from a pre- to post-edited section (Walker & Federici, 2018). These metrics compare the number of altered words between the machine-translated version and the post-edited version of a particular sentence, and so partially represent the technical effort. By analyzing certain edit actions, one may acquire more information about the technical effort.

Blain et al. (2011) comparing statistical and rule-based systems, categorized these activities as post-editing procedures on a language level and established the following PEA classification based on the previous mistake classifications (Blain et al., 2011). The typology of PEA is presented in Figure 2 below:



Noun-Phrase (NP) — related to lexical changes.	• Choice of noun meaning — a noun is changed by another noun changing its meaning • Noun stylistic change — a noun is replaced by a synonym (no meaning change) • Noun number change • Case change • Adjective choice — change in adjective choice for better fit with modified noun • Multi-word change — multiword expression change (meaning change) • NP structure change — structure change of NP • Determiner choice — change in determiner
Verbal-Phrase (VP) — related to grammatical changes	• Verb agreement — correction of agreement in verb • Verb phrase structure change • Verb meaning choice — a verb is replaced by another verb changing its meaning • Verb stylistic change — a verb is replaced by a synonym
Preposition change	
Co-reference change	generally through introduction/removal of a pronoun, or change of a definite to possessive determiner
Reordering	repositioning of a constituent at a better location (adjective, adverb)
PE Error	Post-editor made an error in the review
Misc style	unnecessary stylistic change
Misc	all PEAs that we cannot classify

Fig. 2 Blain et al.’s PEA typology (2011)

Furthermore, the reseachers established the concept PEA as a set of logical edits applied to Post-Editing (PE), in contrast to mechanical edits. A PEA is „minimal“ as long as there isn't a smaller independent edit available. A PEA is referred to as "logical" if the language transition it describes makes sense. Through the introduction of PEA, their study showed that a large part of the PE effort could be classified and automatically learned (Blain et al., 2011).

Koponen (2012) contrasted perceived technical PE effort to real technical effort (as defined by the TER metric), enumerating the sorts of changes for which the discrepancy between perceived and actual PE effort was substantial. Analyzing the data by parts of speech (POS) may imply that overall, modifications involving nouns, verbs, and adjectives require more work than edits involving other POS, as the greatest correlations in both sets mostly involved nouns, verbs, and adjectives. In both groups, phrases with low manual scores had more modified verbs, and in the low TER set, verb matches exhibited one of the greatest associations. On the other hand, there appeared to be particularly substantial relationships between noun-related alterations and the high TER set (Koponen, 2012). In one of the studies presented in Wisniewski et al. (2013), an automatic analysis of post-edits based on Levenshtein distance is performed with only the basic level of substitutions, deletions, insertions, and TER shifts taken into account. These edit actions are analyzed at the lexical level to identify the most often impacted terms (Wisniewski et al., 2013).

Other reseachers Popović et al. (2014) present the technical effort by following five types of edit operations: correction of word form; correction of word order; adding omission; deletion of addition and correction of lexical choice. For error analysis, the completed edit operations are categorised at the word level using the Hjerson automated tool. The output of the post-edited translation was utilized as a reference translation, and the findings are provided as raw counts and edit rates for each category. The edit rate is defined as the ratio of the number of altered words to the total number of words, or sentence length, in the translation output (Popović et al., 2014).

The research of Cui et al. (2023) recruited 33 Chinese postgraduate students concentrating in Translation Studies, comprising 26 females and 7 males between the ages of 22 and 28. They were not yet professional translators, but they may be considered “semiprofessionals“. The Gazepoint GP3 HD Desktop Eye Tracker was used for the experiment due to its efficacy and dependability in data collection, and Translog-II was used to log keystrokes. The results indicated that participants edited more in HT than in PE, indicating that their technical effort was greater in HT. This finding leads to the conclusion that some translators despise PE and prompts us to ponder why translators would dislike a tool that may increase their productivity and translation quality. One theory is that reduced temporal and technical effort does not always equate to decreased cognitive effort, and translators may still endure cognitive effort even when they work quicker and produce translations of higher quality (Cui et al., 2023).

Overall, the assessment of post-editing efforts can be conducted through various techniques and instruments. Several prevalent techniques employed to quantify post-editing efforts encompass time tracking, counts of words, error rates and quality scores. Computer-Assisted Translation (CAT) tools, namely SDL Trados, MemoQ, Wordfast, Smartcat are viable options for measuring post-editing efforts. The aforementioned tools have the capability to monitor post-editing duration, word tallies, and mistake frequencies, in addition to furnishing quality evaluations grounded on diverse metrics.The selection of techniques and instruments for assessing and appraising post-editing endeavors will be contingent upon the particular requirements and objectives of the undertaking, along with the accessible resources.

1.3.4. Post-editor’s profile

With the advent of new technology and instruments, the translator's function has undergone a profound transformation. Some scholars and providers of language services wonder if a translator is still a translator, as it appears that he is becoming only a post-editor, while others question why many translators resist post-editing. Utilizing Bourdieu’s theoretical framework, Sakamoto (2019) describes the social statuses of post-editors and translators in the area of translation. Using the notion of capital, the researcher has been able to comprehend the fight over position-taking between these two distinct sets of stakeholders and why translators are resistant to the move to the new MTPE model. The finding of the study reveals a possible approach to alleviating their resistance: the pay scale of post-editors. Furthermore, the author states that posting is a relatively new activity within the language service sector, and its standing in the translation area is not yet established. In addition, there is a culture of silence regarding the use of MT in the translation process among practitioners. Hence, the evidence supporting the current topic is still limited and in flux (Sakamoto, 2019).

The fact that the post-editor for machine translation must be capable of maintaining the quality of human translation implies that translator’s competence is still a vital element, and it is the most important factor that should be considered when developing the skill set for machine translation post-editors. The main three duties in post-editing include revision of post-edited machine translation output by comparing to the source text, quality control and text checking, and proofreading of post-edited output (Povilaitienė and Kasperė, 2022). O’Brien (2002) in her study claims that post-editing abilities should be taught and presents the reasons why it has be to be done:

- It would help meet the growing demand for translation and faster production times.
- Post-editing skills are distinct from translation skills, and it cannot be assumed that a qualified translator will also be a successful post-editor.
- Graduates would be “comfortable“ with post-editing and better prepared to be productive in a machine translation environment.
- It could increase the adoption of machine translation (O’Brien, 2002).

The more language service companies include machine translation MTPE into their workflows, the greater the need to understand how the PE process is conducted, who the post-editors are, and what abilities they should possess. It is obvious that post-editors need not just fluency in both the source and target languages but also a familiarity with the fundamentals of language, terminology, technology, and IT. International Standard for Translation Services – post editing of machine translation output – Requirements (ISO 18587:2017) claims that the translation service provider (TSP) is responsible for ensuring that the post-editor always accomplishes the following objectives, which are as follows: comprehensibility of the output after editing, correspondence between the content of the source language and the content of the destination language, and compliance with the rules and specifications provided by the TSP for the post-editing process. During post-editing TM output, the TSP is additionally responsible for ensuring that the following standards are satisfied and concerns are taken into consideration:

- Consistency in terminological and lexical use as well as conformity with terminology specific to the domain.
- The use of the conventional syntax, spelling, punctuation, diacritics, special symbols, and abbreviations, as well as other orthographical standards of the target language.
- Conformity with any and all applicable standards.
- Formatted in the appropriate manner.
- Appropriateness for the intended readership in light of the content's intended purpose in the target language.
- Adherence to the terms of the agreement between the client and the TSP (ISO 18587:, 2017).

According to a study conducted by Clara Ginovart Cid et al. (2020), the most crucial primary-level skills for professional poster editors are as follows: the skills required for post-editing machine translation (PEMT) can be categorized into eight distinct areas. These include the ability to perform full PE to achieve human-level quality, the ability to adhere to PE guidelines, the capacity to discern when to work on a segment or discard it, the ability to identify errors in MT output, the ability to perform light PE to achieve satisfactory quality, the ability to apply appropriate correction strategies, the ability to explore new technologies, and the ability to provide recommendations on when PEMT is appropriate for a given text. In fact, the research revealed that, from a human resources perspective, the post-editor's profile is quite similar, if not identical, to the translator's profile. It cannot be expected of a post-editor candidate to provide quality work if translation abilities are not properly honed, and in particular if subject knowledge and revision skills are not maintained. While the industry appears to have a very practical method for determining whether a candidate is a good fit for a position (subject field and CAT tools knowledge were the second- and third-most valued criteria), MTPE training courses are currently more focused on PE skills proper, such as determining when to edit or discard a segment (Cid et al., 2020).

The close relationship between translation and post-editing necessitates that professional post-editors possess a diverse set of skills and knowledge. Post-editing is a commonly employed technique in parallel with machine translation to enhance the effectiveness as well as accuracy of the translation procedure. It is a process that enables human editors to modify machine-translated output, thereby ensuring quality while saving time and resources as opposed to beginning from scratch with an entirely novel translation.

2. Human performance in Translog-II

The present chapter presents the outcomes of a study that utilized the key logging technique and offers a qualitative and quantitative comparative analysis. The aim of the study is to assess machine translation post-editing efforts from EN to LT languages. For this reason, the objectives are to identify the primary problems and main errors according to Blain's typology encountered during post-editing of two different types of content texts, to compare the results to determine the post-editing tactics used in translation processes. In order to proceed, it is imperative to conduct a review of the practices employed by post-editors. This analysis should aim to identify the factors that influence their decision to engage in or abstain from machine translation post-editing (MTPE), the MT tools that are most commonly utilized, and the extent to which machine translation is employed in their daily tasks. This study is intended to analyze the quality of two machine translation outputs in order to determine whether there is a correlation between the level of cognitive effort and the quality of the MT output.

2.1. Methodology of the research

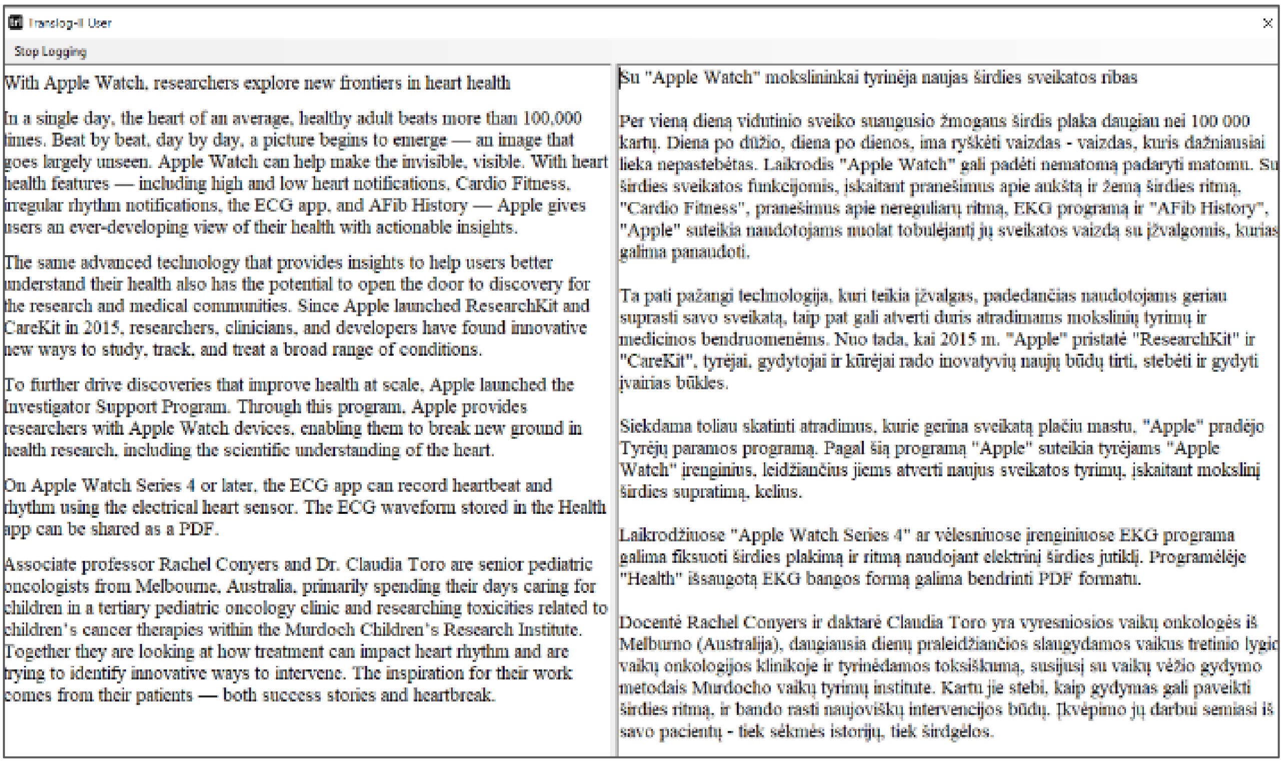
This section of the project is devoted to outlining the methodology employed. The study involved the collection and analysis of data pertaining to the keyboard activity of participants during two post-editing tasks, utilizing the Translog-II software. The quantitative and qualitative comparative analysis was employed for the analysis of the results.

The post-editing tasks were performed by seven first and second level Master's Degree students in Translation Studies from Kaunas University of Technology, who were of Lithuanian nationality. The study's sample comprised individuals with varying levels of expertise in translation, including professional translators, translation students with formal education in the field of translation. Firstly, before doing the task a survey was conducted to determine users experience of post-editing machine translation and can be found in Appendix 1 and Appendix 2. DeepL and Google Translate are the two most commonly utilized machine translation tools.

Secondly, the participants were informed about Translog-II tool which would record their post-editing work and that is a Windows-based software used to record and analyze computer reading and writing operations. Furthermore, they were told that the tool could be used for other computer-based reading or writing tasks as well as for studying translation processes and it contains two main components the Translog-User and the Translog-Supervisor, two interconnected programs (Carl, 2012). They had to post-edit two machine translation outputs at their own pace, without any temporal restrictions and were allowed to use any form of translation assistance or navigating away from the Translog-II interface. This was done for the enhancing of post-editing quality.

Two distinct contextual texts were chosen for the purpose of the study in order to provide a rationale for exploring the specific linguistic results. The focus of the study pertains to the important features of a comparison error analysis. For the machine translation output Deepl tool was selected as according to Cambedda et al's findings (2021), it provides a generally better overall translation performance, especially evident in its rendering of the context and the syntactical structure of the text at the sentence level (Cambedda et al., 2021). The texts underwent machine translation employing the DeepL platform. The software Translog-II was utilized to acquire key-logging data. Figure 3 depicts the screen layout for post-editing activities performed with the Translog-II User interface.

The source text was on the left, while target machine text output was on the right. The target window on the right was chosen mostly because of the vertical layout of CAT tools like Trados and Memoq.



With Apple Watch, researchers explore new frontiers in heart health

In a single day, the heart of an average, healthy adult beats more than 100,000 times. Beat by beat, day by day, a picture begins to emerge — an image that goes largely unseen. Apple Watch can help make the invisible, visible. With heart health features — including high and low heart notifications, Cardio Fitness, irregular rhythm notifications, the ECG app, and AFib History — Apple gives users an ever-developing view of their health with actionable insights.

The same advanced technology that provides insights to help users better understand their health also has the potential to open the door to discovery for the research and medical communities. Since Apple launched ResearchKit and CareKit in 2015, researchers, clinicians, and developers have found innovative new ways to study, track, and treat a broad range of conditions.

To further drive discoveries that improve health at scale, Apple launched the Investigator Support Program. Through this program, Apple provides researchers with Apple Watch devices, enabling them to break new ground in health research, including the scientific understanding of the heart.

On Apple Watch Series 4 or later, the ECG app can record heartbeat and rhythm using the electrical heart sensor. The ECG waveform stored in the Health app can be shared as a PDF.

Associate professor Rachel Conyers and Dr. Claudia Toro are senior pediatric oncologists from Melbourne, Australia, primarily spending their days caring for children in a tertiary pediatric oncology clinic and researching toxicities related to children's cancer therapies within the Murdoch Children's Research Institute. Together they are looking at how treatment can impact heart rhythm and are trying to identify innovative ways to intervene. The inspiration for their work comes from their patients — both success stories and heartbreak.

Su "Apple Watch" mokslininkai tyrinėja naujas širdies sveikatos ribas

Per vieną dieną vidutinio sveiko suaugusio žmogaus širdis plaka daugiau nei 100 000 kartų. Diena po dūžio, diena po dienos, ima ryškėti vaizdas - vaizdas, kuris dažniausiai lieka nepastebėtas. Laikrodis "Apple Watch" gali padėti nematomą padaryti matomu. Su širdies sveikatos funkcijomis, įskaitant pranešimus apie aukštą ir žemą širdies ritmą, "Cardio Fitness", pranešimus apie nereguliarų ritmą, EKG programą ir "AFib History", "Apple" suteikia naudotojams nuolat tobulėjantį jų sveikatos vaizdą su įžvalgomis, kurias galima panaudoti.

Ta pati pažangi technologija, kuri teikia įžvalgas, padedančias naudotojams geriau suprasti savo sveikatą, taip pat gali atverti duris atradimams mokslinių tyrimų ir medicinos bendruomenėms. Nuo tada, kai 2015 m. "Apple" pristatė "ResearchKit" ir "CareKit", tyrėjai, gydytojai ir kūrėjai rado inovatyvių naujų būdų tirti, stebėti ir gydyti įvairias būkles.

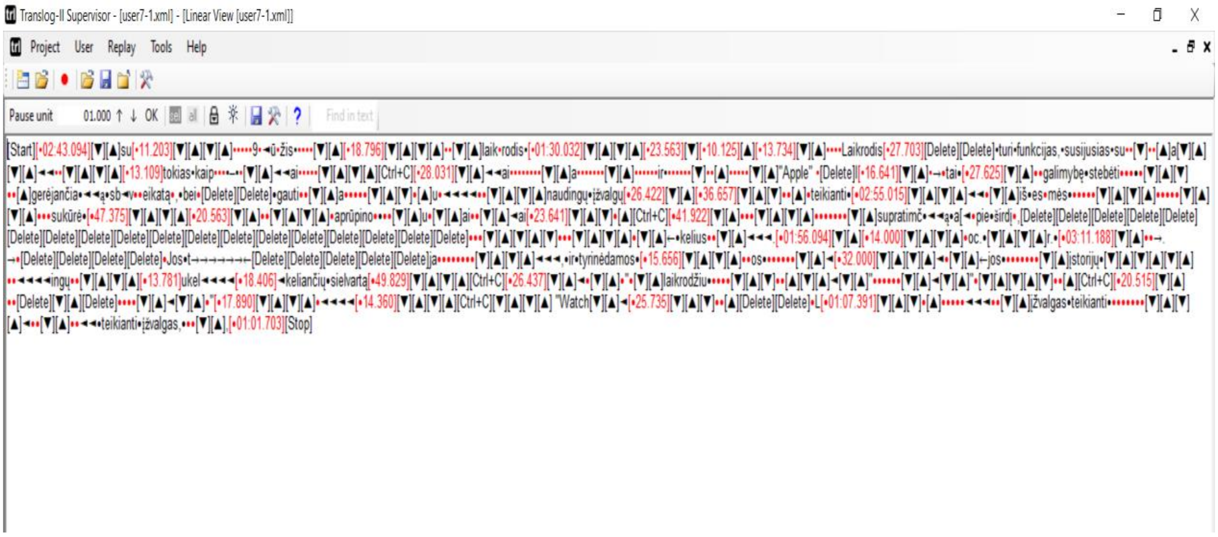
Siekdama toliau skatinti atradimus, kurie gerina sveikatą plačiu mastu, "Apple" pradėjo Tyrėjų paramos programą. Pagal šią programą "Apple" suteikia tyrėjams "Apple Watch" įrenginius, leidžiančius jiems atverti naujus sveikatos tyrimų, įskaitant mokslinį širdies supratimą, kelius.

Laikrodžiuose "Apple Watch Series 4" ar vėlesniuose įrenginiuose EKG programa galima fiksuoti širdies plakimą ir ritmą naudojant elektrinį širdies jutiklį. Programėlėje "Health" išsaugotą EKG bangos formą galima bendrinti PDF formatu.

Docentė Rachel Conyers ir daktarė Claudia Toro yra vyresniosios vaikų onkologės iš Melburno (Australija), daugiausia dienų praleidžiančios slaugydamos vaikus tretinio lygio vaikų onkologijos klinikoje ir tyrinėdamos toksiškumą, susijusį su vaikų vėžio gydymo metodais Murdocho vaikų tyrimų institute. Kartu jie stebi, kaip gydymas gali paveikti širdies ritmą, ir bando rasti naujoviškų intervencijos būdų. Įkvėpimo jų darbui semiasi iš savo pacientų - tiek sėkmės istorijų, tiek širdgėlos.

Fig. 3 Screenshot of the post-editing user’s window of Translog-II

Translog-II tool performs three primary functions: 1) creating a project, 2) running and recording a Translog-II session, 3) performing an analysis of a recorded log file. The final feature aids in User Activity Data (UAD) analysis via linear view: plots a textual representation of UAD; statistics: figures regarding text production, removal, and navigation events; user view: replays the translation session in time; pause plot: depicts in 2D how the text develops over time. Figures 4-6 present visualization of the above mentioned functions. The UAD in the Linear View is represented in a textual manner. The Linear View depicts every instance of key and mouse activity, and any interruptions are denoted by dots, accompanied by a numerical value that signifies the duration between consecutive activities. The resolution of the pause indicator can be adjusted, ranging from 1 millisecond to a user-defined duration. This feature enables the user to obtain a comprehensive understanding of the general temporal organization of a translation session, while minimizing the amount of temporal data. Alternatively, it allows for a closer examination of pausing patterns at a micro-level, with a resolution as fine as a few hundred seconds (Carl, 2012).


Translog-II Supervisor - [user7-1.xml] - [Linear View [user7-1.xml]]
Project
User
Replay
Tools
Help
Pause unit
01.000
OK
Find in text

```
[Start][•02:43.094][▼][▲]su[•11.203][▼][▲][▼][▲]•••••9•◄ū•žis•••••[▼][▲][•18.796][▼][▲][▼][▲]••[▼][▲]laik•rodis•[•01:30.032][▼][▲][▼][▲][•23.563][▼][•10.125][▲][•13.734][▼][▲]••••Laikrodis[•27.703][Delete][Delete]•turi•funkcijas,•susijusias•su••[▼]••[▲]a[▼][▲]
[▼][▲]◄◄••[▼][▲][▼][▲][•13.109]tokias•kaip•••–•[▼][▲]◄◄ai•••••[▼][▲][▼][▲][Ctrl+C][•28.031][▼][▲]◄◄ai••••••••[▼][▲]a•••••••[▼][▲]••••••ir•••••••[▼]••[▲]•••••[▼][▲]"Apple" •[Delete]I[•16.641][▼][▲]•→•tai•[•27.625][▼][▲]••galimybę•stebėti•••••[▼][▲][▼]
••[▲]gerėjančia•◄◄ą•sb◄v••eikatą•,•bei•[Delete][Delete]•gauti••[▼][▲]a•••••[▼][▲][▼]•[▲]u•◄◄◄◄◄••[▼][▲][▼][▲]naudingų•įžvalgų[•26.422][▼][▲][•36.657][▼][▲][▼]••[▲]•teikianti•[•02:55.015][▼][▲][▼][▲]◄◄•[▼][▲]iš•es•mės••••••[▼][▲][▼][▲]•••••[▼][▲]
[▼][▲]•••sukūrė•[•47.375][▼][▲][▼][▲][•20.563][▼][▲]••[▼][▲][▼][▲]•aprūpino••••[▼][▲]u•[▼][▲]ai••[▼][▲]◄ai[•23.641][▼][▲][▼]•[▲][Ctrl+C][•41.922][▼][▲]•••[▼][▲][▼][▲]•••••••[▼][▲]supratimč•◄◄ą•a[◄•pie•širdį•,[Delete][Delete][Delete][Delete][Delete]
[Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete][Delete]•••[▼][▲][▼][▲][▼]•••[▼][▲][▼][▲]•[▼][▲]←•kelius••[▼][▲]◄◄◄.[•01:56.094][▼][▲][•14.000][▼][▲][▼][▲]•oc.•[▼][▲][▼][▲]r.•[•03:11.188][▼][▲]••→.
→•[Delete][Delete][Delete][Delete]•Jos•t→→→→→→←[Delete][Delete][Delete][Delete][Delete]ja••••••••[▼][▲][▼][▲]◄◄◄,•ir•tyrinėdamos•[•15.656][▼][▲][▼][▲]••os•••••••[▼][▲]◄[•32.000][▼][▲][▼][▲]◄•[▼][▲]←jos••••••••[▼][▲]įstorijų•[▼][▲][▼][▲][▼][▲]
••◄◄◄◄ingų••[▼][▲][▼][▲][•13.781]ukel◄◄◄◄[•18.406]◄keliančių•sielvartą[•49.829][▼][▲][▼][▲][Ctrl+C][•26.437][▼][▲]◄•[▼][▲]•"•[▼][▲]laikrodžiu•••••[▼][▲][▼]••[▲][▼][▲]◄[▼][▲]"••••••[▼][▲]◄[▼][▲]"•[▼][▲][▼][▲][▼]••[▲][Ctrl+C][•20.515][▼][▲]
••[Delete][▼][▲][Delete]••••[▼][▲]◄[▼][▲]•"[•17.890][▼][▲][▼][▲]•◄◄◄◄[•14.360][▼][▲][▼][▲][Ctrl+C][▼][▲][▼][▲] "Watch[▼][▲]◄[•25.735][▼][▲][▼]••[▲][Delete][Delete]•L[•01:07.391][▼][▲][▼]•[▲]•••••◄◄◄••[▼][▲]įžvalgas•teikianti••••••••[▼][▲][▼]
[▲]◄••[▼][▲]••◄◄•teikianti•įžvalgas,•••[▼][▲].[•01:01.703][Stop]
```

Fig. 4 Screenshot of Linear view of Translog-II supervisor component

The red dots stand for one second pause made by user while black arrows up and down denote the presence of mouse activity and represent the act of selecting a specific location on the display, typically to initiate the process of produced writing. The specification of keyboard activities is denoted by square brackets, wherein the act of pressing the DELETE button is represented by [DELETE], and the activity of using CONTROL and C at the same time for copying is represented by [CTRL+C].

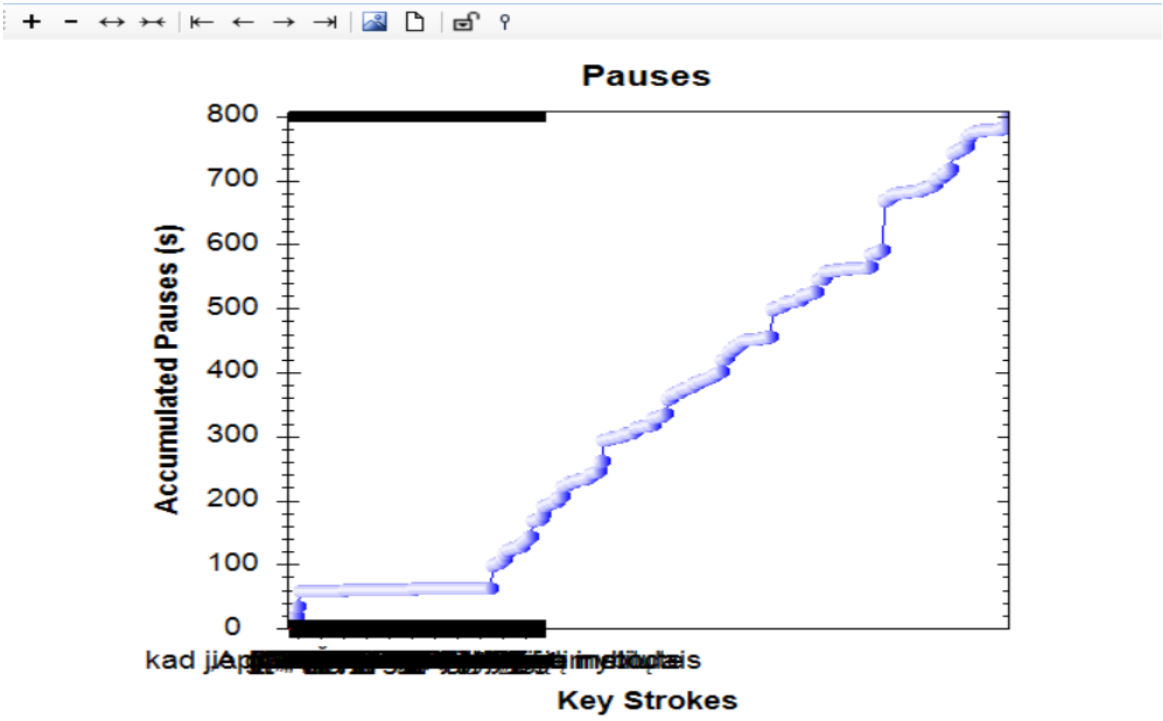
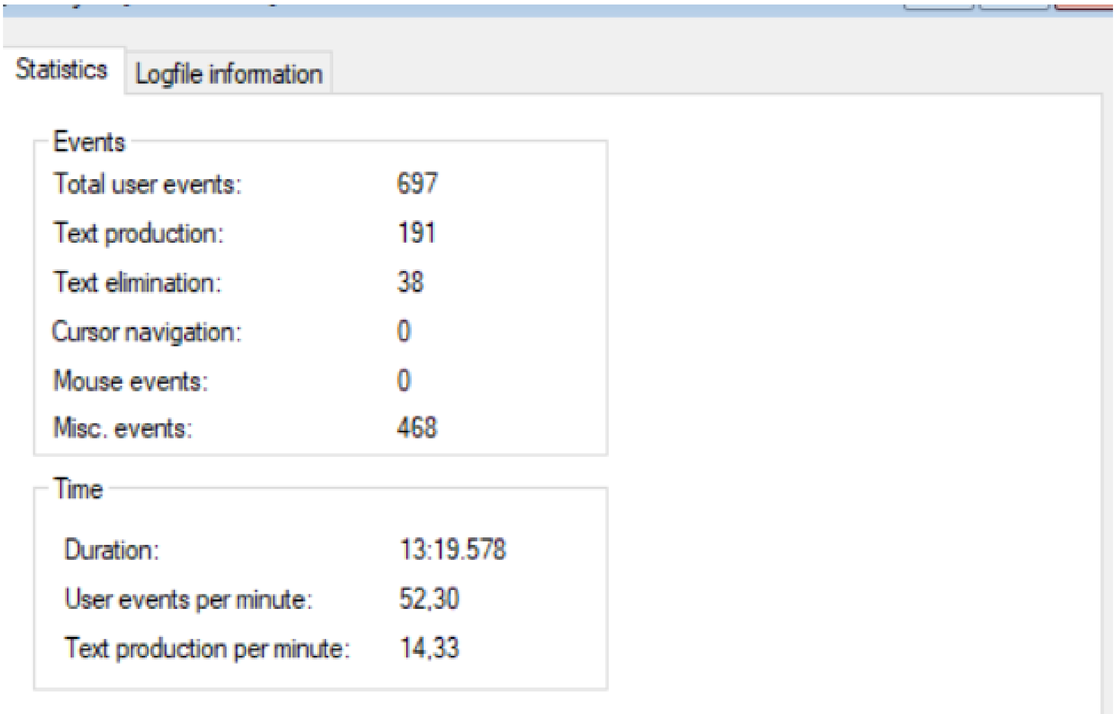


Fig. 5 Screenshot of Pause Plot view of Translog-II supervisor component

The Pause Plot illustrates the advancement of produced text. The tool defines the user's periods of activity and inactivity. The blue dots represent a specific action, which can either be the act of gazing or typing.The horizontal display of graphemes is accompanied by a vertical line that indicates the length of the participant's pauses. The examination of the Pause Plot bears resemblance to the analysis of Linear View in that it portrays pauses in a two-dimensional format.



Events

Total user events: 697

Text production: 191

Text elimination: 38

Cursor navigation: 0

Mouse events: 0

Misc. events: 468

Time

Duration: 13:19.578

User events per minute: 52,30

Text production per minute: 14,33

Fig. 6 Screenshot of Statistics view of Translog-II supervisor component

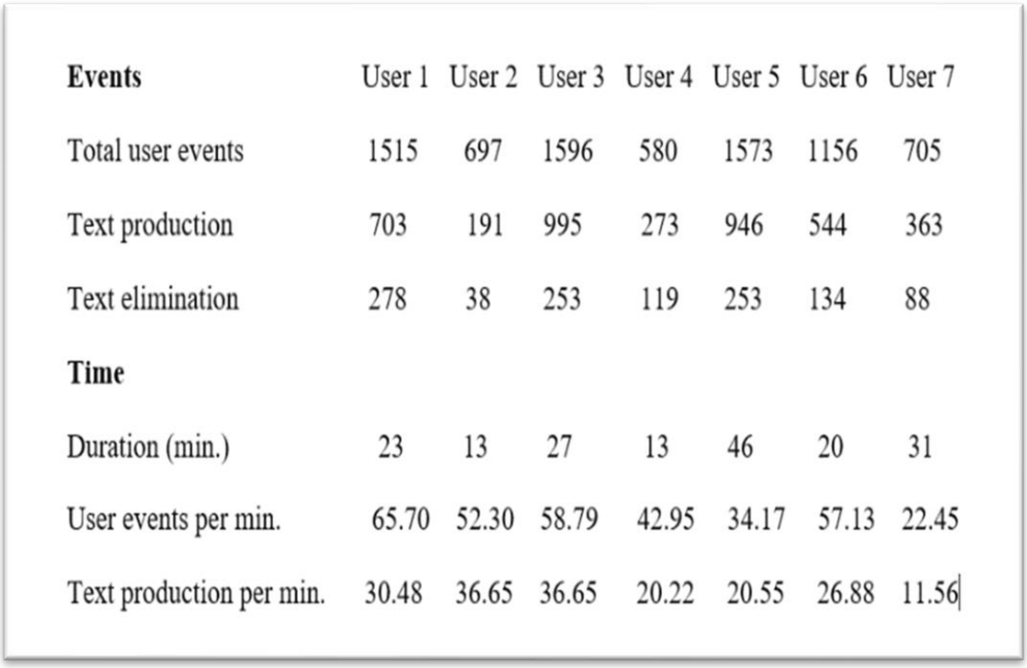
The field of statistics comprises two primary sections, each containing several sub-sections. The initial segment documents the total number of user events, textual production, and deletions, among other factors. Conversely, the subsequent section pertains to the duration of post-editing.

2.2. Analysis of the Translog-II digital data of human post-editing processes

The data obtained from the experiment is subjected to quantitative, qualitative comparative analysis. The data related to each text was collected, quantified, and will be presented and analyzed subsequently. The Statistics function was utilized to obtain the outcomes, which revealed a total amount of user events and the duration of post-editing.

2.2.1. Statistics

Through the analysis of the duration of post-editing and the types of modifications implemented by post-editors, it is feasible to deduce the level of proficiency of the participants. Talking about the survey results and posteditor's experinece it must be mentioned that there were two participants who reported having less than one year of experience, three participants with experience ranging from one to three years, and two participants with three or more years of experience. Out of the total of six interviewers, only one of them refrains from utilizing machine translation tools. Out of six respondents, three reported using them rarely, two reported using them usually, and one reported always using them. Five participants employed the post-editing technique, while two participants refrained from using it. The primary challenges which were indicated in the survey by participants were that machine translation tools includes inaccurate noun case usage, incorrect abbreviation usage, literal translation, inadequate use of terminology, and inappropriate translation of complex sentence structures. The data analysis is conducted utilizing the Statistics function produced by Translog-II. As previously stated, the present text is a research article sourced from the website of Apple company, comprising a total of 289 word. The total results are presented in Figure 7:

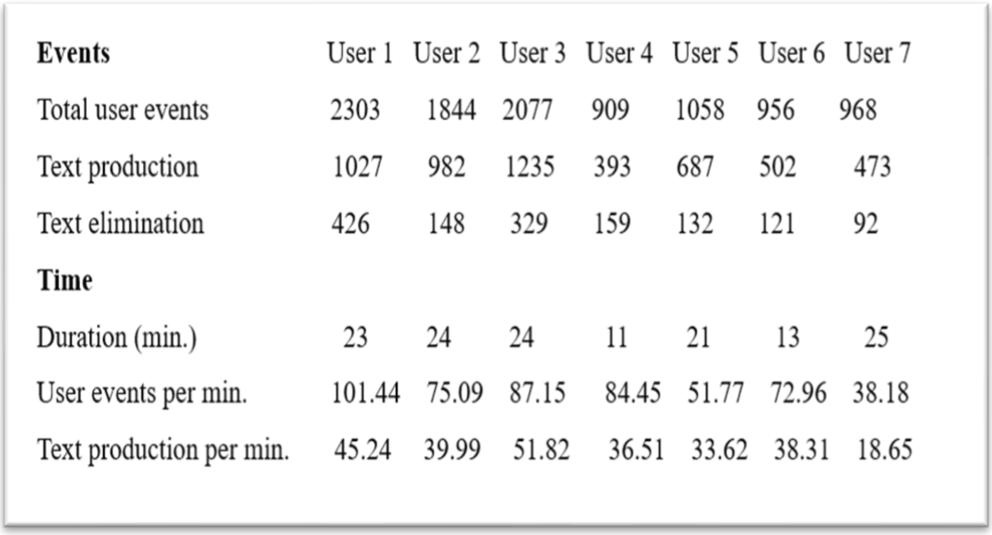


Events	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Total user events	1515	697	1596	580	1573	1156	705
Text production	703	191	995	273	946	544	363
Text elimination	278	38	253	119	253	134	88
Time							
Duration (min.)	23	13	27	13	46	20	31
User events per min.	65.70	52.30	58.79	42.95	34.17	57.13	22.45
Text production per min.	30.48	36.65	36.65	20.22	20.55	26.88	11.56

Fig. 7 Results of "Apple" website research article based on Translog-II Statistics function

A total of four participants engaged in over 1000 activities, while three other participants completed a range of 580-705 events during the post-editing phase of the text. User 3 exhibited the highest number of user and text production events in total. The mean duration of post-editing is 24.7 minutes. The users engaged in post-editing activities for varying durations. Specifically, the shortest duration was 13 minutes, while the longest duration was 46 minutes, attributed to User 5. Hence, it is evident that during the process of text production, User 5 exhibited a typing speed of 205 characters per minute, which was marginally higher than the slowest recorded speed of User 4, who typed at a rate of 20.22 characters per minute. Furthermore, it has been observed that User 5 conducted the most extensive post-editing, ranked second in terms of text elimination, with a total of 253 events. According to the survey analysis of User 5, it can be inferred that the participant's level of experience is below one year. Furthermore, the respondent indicates that he does not employ machine translation tools in his professional translation practice, which has an impact on the extended duration of post-editing time required to achieve satisfactory results. User 1 achieved the top rank in text elimination with a score of 278 and exhibited an impressive typing speed of 65.70 events per minute. The average post-editing time suggests that User 1 may have certain level of experience. The duration of post-editing undertaken by users 2 and 4 was identical. However, it is noteworthy that user 4 removed a greater amount of text compared to user 2. Furthermore, the reason for this disparity in performance can be attributed to the fact that User 4, as per the pre-experiment survey, possesses prior experience in post-editing spanning a duration of 1-3 years, whereas User 2 lacks such experience.

The subsequent phase of the examination involves scrutinizing the outcomes obtained during the process of revising a user guide intended for end-users, which spans a total of 282 words. It is usually believed that the post-editing of the machine-translated output of the user guide will be less difficult task due to the presence of numerous imperative verb phrases. Figure 8 displays the outcomes as follows:



Events	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Total user events	2303	1844	2077	909	1058	956	968
Text production	1027	982	1235	393	687	502	473
Text elimination	426	148	329	159	132	121	92
Time							
Duration (min.)	23	24	24	11	21	13	25
User events per min.	101.44	75.09	87.15	84.45	51.77	72.96	38.18
Text production per min.	45.24	39.99	51.82	36.51	33.62	38.31	18.65

Fig. 8 Results of "Apple Watch" user guide based on Translog-II Statistics function

The range of user events encompasses a total of 909 to 2303, whereas in the initial text, it was reported to be between 580 and 1515. The range of text production events observed falls between 393 and 1235, while the range of text elimination events observed falls between 92 and 426. The numerical values exhibit a significant disparity when contrasted with the representation of previous text. It is noteworthy that User 1 primarily engaged in user events and deletion. User 4 exhibited the lowest total number of user events and text production, although not in terms of elimination. In addition, it should be noted that while the time duration may be brief, it does not necessarily correlate with the level of experience possessed by the post-editor. The average duration of post-editing is 20 minutes.

In comparison, the range of user events per minute is between 38.18 and 101.44, and the range of text production is between 18.65 and 51.82. Notably, the similarity between user events per minute is more pronounced than the similarity between text production. User 1 exhibited a higher frequency of 101.44 events per minute, whereas User 7 demonstrated a comparatively lower frequency of 38.18 events per minute. Furthermore, it can be observed that the User 7 has the longest time duration, while the occurrences of text elimination, production, and user events are the least frequent. The presence of pauses during post-editing may suggest that they were caused by either a lack of experience on the part of the editor or the complexity of the text being edited. The latter rationale appears to be more substantiated, as all seven participants exhibited a greater number of events in comparison to the first text.

Upon comparing the two texts, it was observed that the initial text required an additional four minutes for post-editing, despite the Translog-II program recording fewer events. On average, the second text was executed within a duration of 20 minutes, yet encompassing a greater number of recorded events. In conclusion, it is possible that post-editors exhibit greater familiarity with the translation of manual guides. Another factor to consider is that additional modifications were implemented due to inadequate machine translation results.

2.2.2. Pause Plot

The Pause Plot graphically represents the progress of generated produced text. The aforementioned tool establishes the user's intervals of reactions and pausing. The blue dots are indicative of a particular action, namely either the act of gazing or typing.The graphemes are arranged horizontally and are accompanied by a vertical line that denotes the duration of the participant's pauses. The examination of the Pause Plot exhibits similarities to the evaluation of the Linear View in that it represents pauses in a bi-dimensional configuration. Nevertheless, it is noteworthy to acknowledge that its informative worth is relatively inferior to that of data obtained through Linear View. As a result, the investigation lacks a comprehensive analysis of the aforementioned subject matter and instead presents only limited observations. Figures 9-10 illustrate and indicate that the red arrows serve as markers for interruptions initiated by individuals. A couple of random visual representations are displayed here:

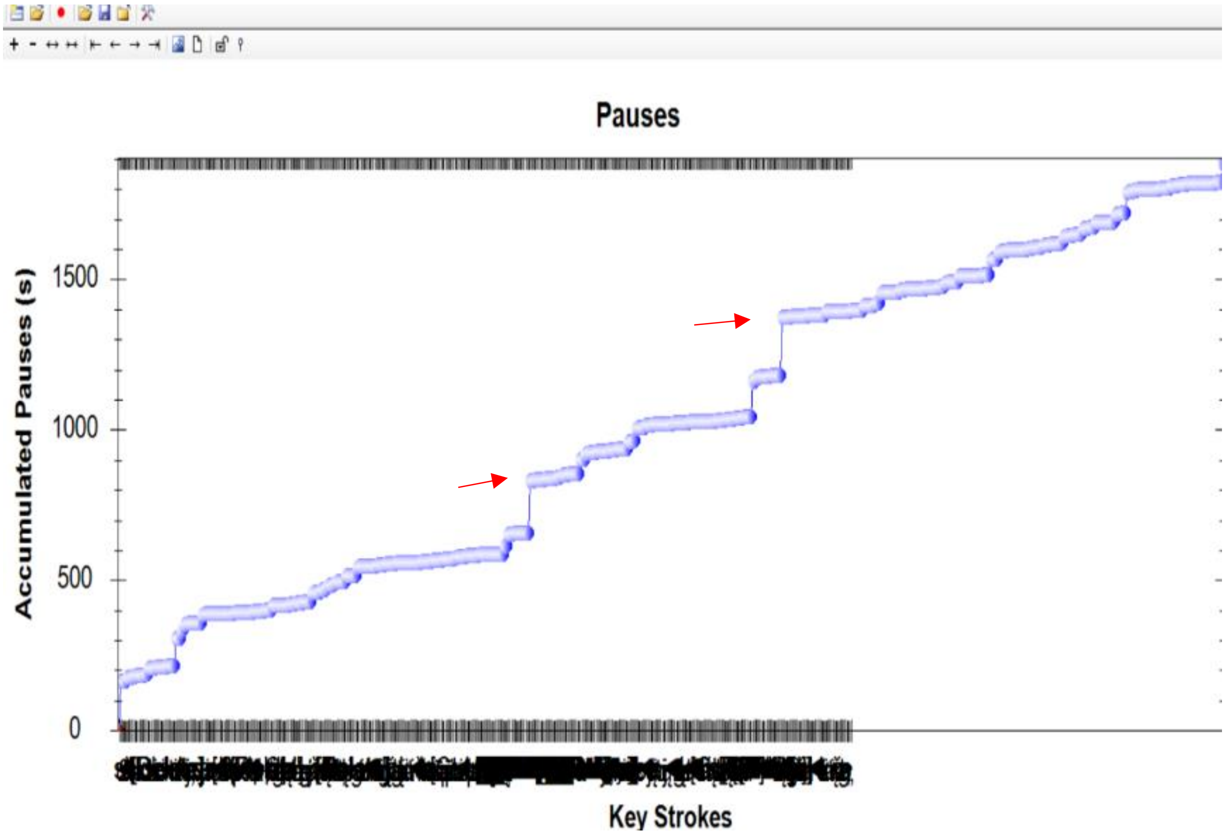


Fig. 9 Pause Plot grapheme (User 7)

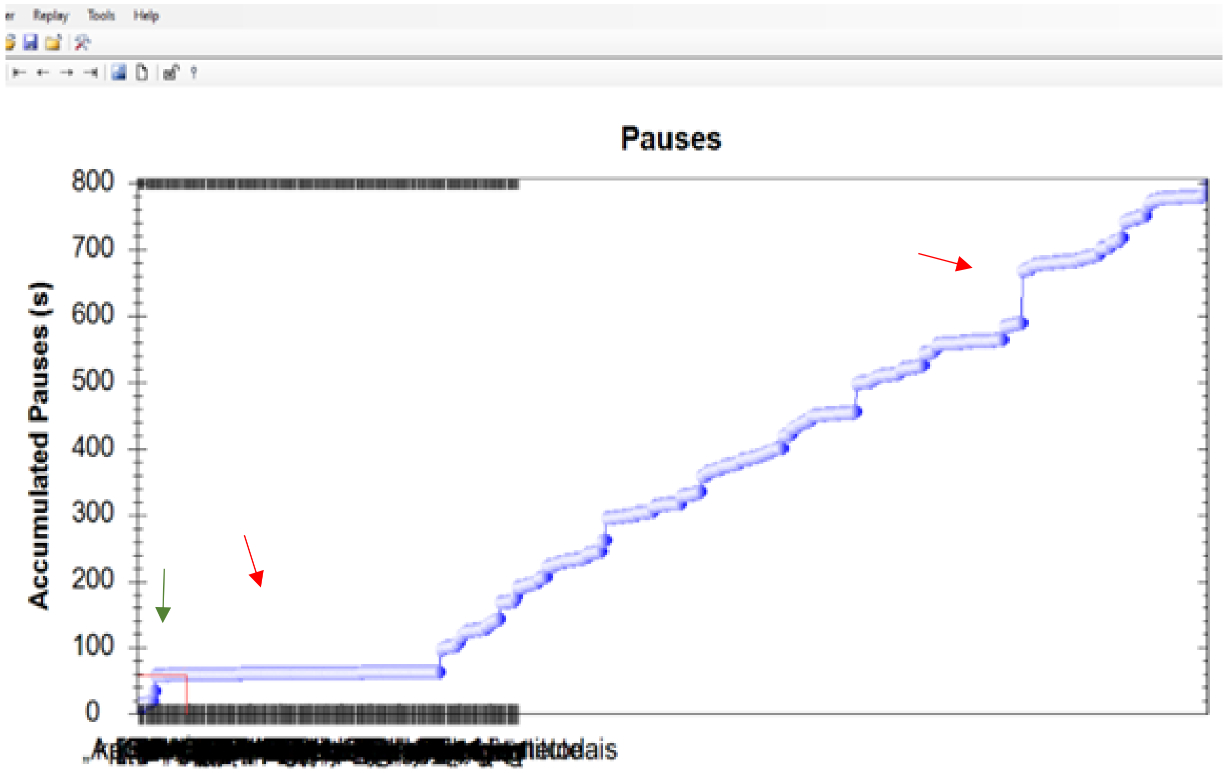


Fig. 10 Pause Plot grapheme (User 2)

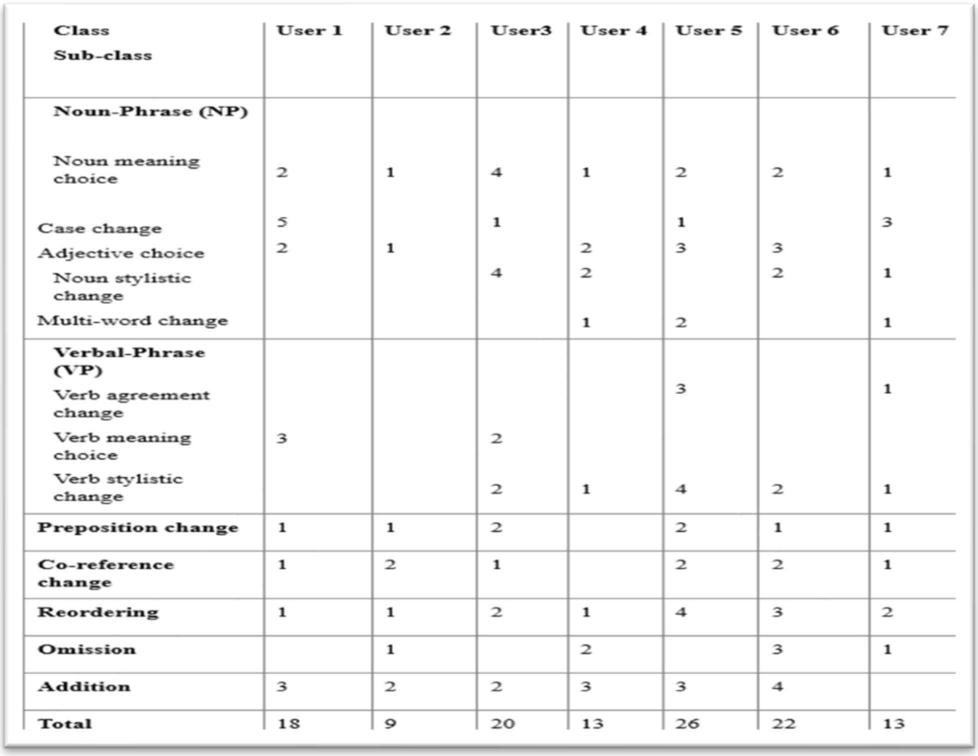
Upon comparing the two graphemes, it can be posited that User 2 engaged in certain activities at the onset of the task (refer to Figure 10, wherein the green arrow denotes reactions), followed by a pause exceeding two minutes, indicating a state of hesitation. At the beginning, User 7 demonstrated promptness in their approach to the post-editing task. However, during the progress of the task, they exhibited intermittent pauses. Upon utilizing the replay functionality, it was observed that User 2 ceased involvement after the initial sentence of the assigned task.

2.3. Analysis of the post-editing actions according to Blain's typology

The data review was conducted utilizing the Translog-II replay function. To conduct an experimental study, two texts were selected in order to determine whether text content heterogeneity and complexity affect post-editing quality. Seven participants performed two tasks, which then underwent analysis using Blain et al.'s (2011) typology. As it was noted in literarure review, the researchers conducted a study that compared statistical and rule-based systems. They identified post-editing procedures on a language level and developed a PEA classification based on mistake classifications. The present research is based on Blain et al.'s (2011) typology but minor modifications have been implemented, specifically the exclusion of the determiner sub-class due to the absence of determiners in the Lithuanian language. Additionally, the inclusion of omission and addition classes, previously utilized in the PEA, has been incorporated. Moreover, it has been observed that post-editing actions relate to multiple sub-classes.

2.3.1. The results of the "Apple" website research article

The research material from the Apple company website was the first assignment that participants post-edited. It was retrieved from the website (https://www.apple.com/newsroom/2023/02/with-apple-watch-researchers-explore-new-frontiers-in-heart-health/). The text has 289 words. The source text in question is provided in Appendix 3. Figure 11 presents the outcomes of the post-edited actions (PEA) conducted on each participant, categorized based on Blain et al.'s (2011) taxonomy. The cumulative number of post-editing activities performed by all post-editors is 121:



Class Sub-class	User 1	User 2	User3	User 4	User 5	User 6	User 7
Noun-Phrase (NP)							
Noun meaning choice	2	1	4	1	2	2	1
Case change	5		1		1		3
Adjective choice	2	1		2	3	3	
Noun stylistic change			4	2		2	1
Multi-word change				1	2		1
Verbal-Phrase (VP)							
Verb agreement change					3		1
Verb meaning choice	3		2				
Verb stylistic change			2	1	4	2	1
Preposition change	1	1	2		2	1	1
Co-reference change	1	2	1		2	2	1
Reordering	1	1	2	1	4	3	2
Omission		1		2		3	1
Addition	3	2	2	3	3	4	
Total	18	9	20	13	26	22	13

Fig. 11 Post-edited actions (PEA) of the "Apple" website research article

The analysis of the results reveals that the majority of the PEA instances produced pertain to alterations in Noun-Phrases (NP) were noun meaning choice. The following is a list of PEA, along with examples for each class and sub-class. In this context, ST denotes the source text, TT denotes machine translation target, and PE1-7 denotes post-editing examples of seven users. Blain et al.'s typology identifies the Noun-Phrase class as the initial category, which comprises five sub-classes. The initial subclass is related to the linguistic category of nouns denoting selection, whereby modifications are effected through the substitution of one noun with another.

Noun meaning choice

Ex 1 ST: [...] *Beat by beat* [...]
TT: [...] *Diena po dūžio* [...]
PE1-6 : [...] *Dūžis po dūžio* [...]

Ex 2 ST: [...] *The inspiration for their work* [...]
TT: [...] *Įkvėpimo jų darbu*i [...]
PE3 *: [...] Ryžtas jų darbui* [...]

Ex 3 ST: [...] *On Apple Watch Series 4 or later* [...]
TT: [...] *Laikrodžiuose "Apple Watch Series 4" ar vėlesniuose įrenginiuose* [...]
PE5 *: [...] Su "Apple Watch Series 4" arba vėlesnio leidimo laikrodžiuose* [...]

The majority of participants in Example 1 resolved the inaccurate machine translation output phrase “diena po dūžio” to “dūžis po dūžio”. One of the participants did not provide the correct response. Example 2 demonstrates that post-editor 3 substituted the noun “įkvėpimas” with the noun ryžtas”, which possesses a slightly divergent connotation. The phrase “ryžtas jų darbui” appears to be ungrammatical in Lithuanian, while the machine-translated sentence was superior, albeit with an incorrect usage of the pronoun “jų”. The third example demonstrates that the fifth post-editor utilized a noun-meaning choice strategy by substituting the term “įrenginiuose”" with the more specific term “laikrodžiai”. This was a judicious decision, as “įrenginiai” is a broad term that can refer to any device.

The Lithuanian language exhibits a diverse system of grammatical cases, comprising a total of seven, in contrast to English, which predominantly employs prepositions. It is a prevalent practice among post-editors to implement alterations in the case of textual content. The subsequent instances exemplify alterations in case. 10 modifications were executed with regard to this specific sub-class.

Case change

Ex 4 ST: [...] *Apple provides researchers with Apple Watch devices* [...]
TT: [...] *“Apple” suteikia tyrėjams “Apple Watch” įrenginius* [...]
PE1 : [...] *“Apple” aprūpina tyrėjus “Apple Watch” įrenginiais* [...]

Ex 5 ST: [...] *Including high and low heart notifications* [...]
TT: [...] *Pranešimus apie nereguliarų ritmą* [...]
PE1 : [...] *Tokias kaip pranešimai apie nereguliarų ritmą* [...]

Example 4 demonstrates a transformation from the term “tyrėjams” to “tyrėjus”, while in example 5, “pranešimus” is altered to “pranešimai”. In both instances, it is evident that alterations were prompted by external modifications. Specifically, in the initial example, the change was of a stylistic nature, while in the latter, it involved the inclusion of multiple words. Upon analyzing the omitted examples, it is noted that the outcome remains consistent.
The other sub-class which was analyzed was adjective choice which belongs to Noun-Phrase class as well. In total there were 11 modifications performed:

Adjective choice

Ex 6 ST: [...] *Innovative ways to intervene* [...]
TT: [...] *Naujoviškų intervencijos būdų* [...]
PE6 : [...] *Inovatyvių intervencijos būdų* [...]

Ex 7 ST: [...] *Developing view of their health* [...]
TT: [...] *Tobulėjantį jų sveikatos vaizdą* [...]
PE4 : [...] *Kintantį jų sveikatos vaizd*ą [...]

Ex 8 ST: [...] *Primarily spending their days* [...]
TT: [...] *Daugiausia dienų praleidžiančios* [...]

PE5: [...] *Dauguma dienų praleidžiančios* [...]

Example 6 illustrates the transformation of the term “naujoviškas” to "inovatyvių”, whereas in example 7, “tobulėjantį” is replaced with “kintantį”. Additionally, example 8 demonstrates the substitution of “daugiausia” with “dauguma”. The sentences underwent primarily stylistic modifications to the adjectives. Example 7 illustrates an instance where the machine output was deemed incorrect due to the use of the phrase “tobulėjantis vaizdas” which is deemed an improper expression in the Lithuanian language. The aforementioned example necessitated the change. With the exception of one, all post-editors implemented modifications relating to the adjective, though applying different strategies. The stylistic change is related to a synonym which replaces noun with no change of meaning. There are 10 modifications made related to stylistic change. Compared to noun meaning choice, the translation technique of noun stylistic modification was less frequently adopted. It can imply that the noun-related output of the machine translation was not sufficient.

Stylistic change of noun

Ex 9 ST: [...] *New frontiers in heart health* [...]
TT: [...] *Atrado naujas širdies sveikatos ribas* [...]
PE1 : [...] *Nustatė naujas širdies ribas sveikatinim*[...]

Ex 10 ST: [...] *Primarily spending their days* [...]
TT: [...] *Daugiausiai dienų* [...]
PE6 : [...] *Daugiausiai laiko* [...]

Ex 11 ST: [...] *Researching toxicities related* [...]
TT: [...] *Tyrinėdamos toksiškumą, susijusį* [...]
PE6 : [...] *Tyrinėdamos nuodingumą. susijusį* [...]

All of the illustrated examples above demonstrate the implementation of stylistic changes through the substitution of a noun with a synonymous noun. The phrase in the example 9 “atrado naujas ribas širdies sveikatinime” could be rephrased “nustatė naujas ribas sveikatos srityje” instead of using the term “sveikatinimas” as it would be grammatically incorrect to use a noun in a locative case. The other subclass that refers to noun phrases is the alteration of multiple words.This technique was employed by three post-editors, resulting in a total of four modifications. The following are two examples that demonstrate the aforementioned alteration:

Multi-word change

Ex 12 ST: [...] *Both success stories and heartbreak* [...]
TT: [...] *Tiek sėkmės istorijų, tiek širdgėlos* [...]
PE7 : [...] *Tiek sėkmingų, tiek keliančių sielvartą* [...]

Ex 13 ST: [...]*To break new ground in health research* [...]
TT: [...] *Atverti naujus sveikatos tyrimų keliu*s [...]
PE5 : [...] *Tyrinėti dar neatrastus sveikatos supratimo kelius* [...]

With respect to example 12, it was observed that the aforementioned phrase remained unaltered in all instances of post-editing, except for the case of post-editor 7. One could make the argument that the post-editor proficiently carried out a modification that involved several words, ultimately producing a phrase that gives the impression of being the norm. A similar assertion can be made regarding example 13, as the machine translation output of “atverti sveikatos tyrimų kelius” is deemed inadequate. The decision to modify the phrase to “tyrinėti dar neatrastus sveikatos supratimo kelius” was an appropriate one.

The second class that was analyzed refers to the Verb-Phrase category, which comprises three sub-classes, namely: verb meaning selection, verb agreement modification, and verb stylistic modification. A total of 19 modifications were implemented. The majority of these changes are related to alterations in verb style, while the minority involve modifications in verb agreement. In total there were only 4 modifications performed related to this strategy. The following examples are enumerated below:

Verb agreement change

Ex 14 ST: [...] *To further drive discoveries that improve* [...]
TT: [...] *Siekdama toliau skatinti atradimus* [...]
PE5 : [...] *Siekiant toliau skatinti sveikatą* [...]

Ex 15 ST: [...] *Using the electrical heart* [...]
TT: [...] *Naudojant elektrinį širdies* [...]
PE7 : [...] *Kuri naudoja elektrinį širdies* [...]

Only two post-editors employed verb agreement alteration. In example 14, the term “siekdama” is replaced with “siekiant” while in example 15, “naudojant” is substituted with “kuri naudoja” through the addition of the conjunction “kuri”. It could also be asserted that the post-editor employed the technique of addition, which is frequently utilized in other instances as well. As exemplified in example 16, it is possible to identify a shift in style as well as the application of “addition” strategy. In the given example, the verb “gali” has been replaced with “suteikia” to enhance the stylistic impression of the phrase. Additionally, the inclusion of the word “galimybę” further enriches the sentence.

Verb stylistic change

Ex 16 ST: [...] *Has the potential to open the do*or [...]
TT: [...] *Taip pat gali atverti duris* [...]
PE5 : [...] *Suteikia galimybę atverti duris* [...]

Ex 17 ST: [...] *Trying to identify* [...]
TT: [...] *Bando rasti naujoviškų* [...]
PE3 : [...] *Bando atrasti naujoviškų* [...]

Verb meaning choice

Ex 18 ST: [...] *To open the door to discovery* [...]
TT: [...] *Atverti duris atradimams* [...]

PE3 : [...] Suteikti naudos *mokslinių* [...]

Ex 19 ST: [...] *Comes from their patients* [...]
TT: [...] Semiasi *iš savo pacientų* [...]
PE1 : [...] Atranda *iš savo pacientų* [...]

The technique of employing a verb meaning choice was not extensively utilized. The overall number of modifications implemented by a subset of two out of seven participants was limited to merely five. A change from the verb “atverti” to “naudoti” has been noted. Additionally, it can be asserted that the post-editor 3 implemented alternative strategies. The other sub-classes are preposition change and co-reference change. Except for one post-editor, all of the edits involved prepositions and co-references. This sub-class underwent 9 changes in total. One post-editor did not make changes related to this technique. The examples are below:

Preposition change

Ex 20 ST: [...]*With Apple Watch* [...]
TT: [...] *Su “Apple Watch”* [...]
PE3 : [...] *Naudojantis “Apple Watch”* [...]
PE4: [...] *Pasitelkdami “Apple Watch”* [...]
PE5: [...] *Laikrodžių pagalba “Apple Watch”* [...]

This particular subclass has undergone a total of nine modifications. A certain post-editor refrained from making modifications pertaining to this particular technique. The machine translation result was incorrect with regard to co-reference alterations for Lithuanian language as reflexive pronoun “savo” instead of “jų” must be used according to lexical rules. Almost all participants changed the pronoun into „savo“ or „šiam“.

Co-reference change

Ex 21 ST: [...] *The inspiration for their work* [...]
TT: [...] *Įkvėpimo jų darbui* [...]
PE5: [...] *Įkvėpimo šiam darbui* [...]
PE6: [...] *Įkvėpimo savo darbui* [...]

The modifications pertained to the removal of the preposition “su” and its replacement with a noun phrase, as demonstrated in example 20 by post-editor 5, or the use of verbal parts of speech such as “pasitelkdami” or “naudojantis”. The application of co-reference change was executed by six users, resulting in a total of nine instances. Notably, user 4 failed to modify the co-reference change, which was attributed to an error on the part of the post-editor. The following examples illustrate an alternative strategy for reordering, specifically a technique in which the post-editor modifies the arrangement of words. Total changes performed in this class is 14:

Reordering

Ex 22 ST: [...] *To open the door to discovery for the research* [...]
TT: [...] *Atverti duris atradimams mokslinių tyrimų* [...]

PE5 : [...] *Atverti duris* naujiems moksliniams tyrimams [...]

Ex 23 ST: [...] *Oncology clinic and researching toxicities related* [...]
TT: [...] *Onkologijos klinikoje ir tyrinėdamos toksiškumą* [...]
PE7 : [...] *Tyrinėdamos su vaikų, sergančių vėžiu, gydymo metodais susijusį nuodingumą* [...]

The class focused on reordering is ranked second in terms of the number of modifications made. One could argue that this strategy is particularly challenging due to the need for accurate noun case changes and the implementation of other appropriate techniques. The given example illustrates significant modifications that necessitate increased temporal and cognitive exertions, as well as a heightened level of expertise. The omission class underwent relatively fewer modifications such as class reordering or addition. The text underwent seven modifications and was reviewed by four post-editors. Examples of the modifications are provided below:

Omission

Ex 24 ST: [...] *A picture begins to emerge — an image that* [...]
TT: [...] *Pradeda ryškėti vaizdas - vaizdas, kuris* [...]
PE3 : [...] *Pradeda ryškėti vaizdas, kuris* [...]

Ex 25 ST: [...] *The inspiration for their work* [...]
TT: [...] *Įkvėpimo jų* darbui [...]
PE7 : [...] *Įkvėpimo darbui* [...]

Regarding these instances of post-editing, it can be asserted that they have been effectively edited, with the alterations made by the post-editors regarded as necessary. The sentence simplicity was improved by removing unnecessary words “jų” and “vaizdas”. The addition method was employed 17 times, which is the most. According to the post-editor's perspective, the substance of some lengthy sentences containing survey participant responses may have been lost. Therefore, it could be argued that the utilization of the “Addition” method primarily aimed to provide a coherent view for the end-user. The following examples are presented below:

Addition

Ex 26 ST: [...] *Has the potential to open the do*or [...]
TT: [...] *Taip pat gali atverti duris* [...]
PE5 : [...] *Suteikia galimybę atverti duris* [...]

Ex 27 ST: [...] *Irregular rhythm notifications* [...]
TT: [...] *Pranešimus apie nereguliarų ritmą* [...]
PE1 : [...] *Pranešimus apie nereguliarų širdies ritmą* [...]

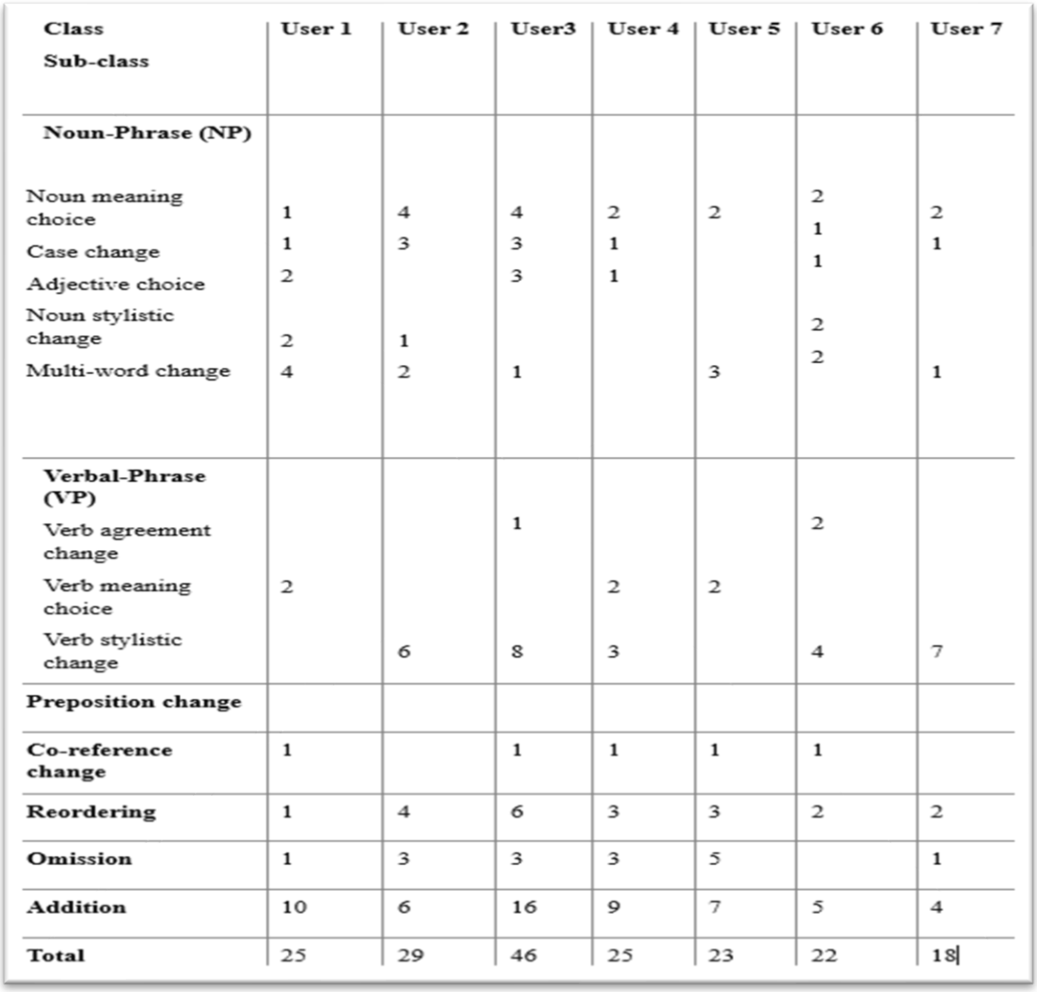
Ex 28 ST: [...] *Since Apple launched* [...]
TT: [...] *Nuo tada, kai 2015 m. “Apple”* [...]
PE1 : [...] *Nuo tada, kai 2015 m.* “*Apple” produkcija* [...]

The implementation of the addition strategy in the context of example 26 appears to have been primarily motivated by stylistic considerations, whereas in the cases of examples 27 and 28, its use appears to have been driven by a desire to enhance clarity. The final instances suggest that the primary terms that were appended were predominantly affixed to the noun “Apple”.

In summary, what concerns noun related changes it is observed that the majority of changes are related to adjective selection and noun meaning change in Noun-Phrase class. When Verbal-Phrase post-editing adjustments are analyzed, it is shown that the majority of the changes were made for stylistic considerations, which had an impact on noun case alterations. Due to the wider variety of noun cases in Lithuanian than in English, verb stylistic, reordering, addition, ommision changes are one of the reason for case shift. The identification of deletions and additions is another interesting result of this study. When comparing the two, it can be seen that the “Addition” approach was used in seventeen occasions and the “Omission” technique in seven. Upon analyzing the results, it was found that the predominant strategy employed was “Addition”, specifically for explanatory purposes. This involved the inclusion of words such as “įmonė”, “kompanija”, and “laikrodis” following the term “Apple”. All participants utilized the “Reordering” technique during post-editing, as the Lithuanian language exhibits a less rigid word order in comparison to English.

2.3.2. The results of the “Apple watch” user guide

The present text, comprising 282 words, has been sourced from the official website of the technology company “Apple”. The selection of this text for post-editing is based on the premise that user manuals and guides typically necessitate minimal post-editing due to their utilization of imperative verbs and concise, unambiguous, and directive sentence structures. However, it is essential that these documents maintain both semantic and grammatical accuracy. Another rationale for selecting it was the significance of the presence of analogous terminology in both texts. The participants were assigned a second task without any specified time constraints. Figure 12 presents post-edited actions (PEA) conducted by seven users, detailing the modifications made to each class or sub-class. The overall quantity of post-editing activities executed by the entirety of post-editors is 188. Noun meaning choice class has undergone 15 modifications, while the case form has undergone 9 changes and the selection of adjectives has been altered 6 times. There were 5 stylistic changes made to the noun usage, and 13 changes made to the use of multi-word expressions.



Class Sub-class	User 1	User 2	User3	User 4	User 5	User 6	User 7
Noun-Phrase (NP)							
Noun meaning choice	1	4	4	2	2	2	2
Case change	1	3	3	1		1	1
Adjective choice	2		3	1		1	
Noun stylistic change	2	1				2	
Multi-word change	4	2	1		3	2	1
Verbal-Phrase (VP)							
Verb agreement change			1			2	
Verb meaning choice	2			2	2		
Verb stylistic change		6	8	3		4	7
Preposition change							
Co-reference change	1		1	1	1	1	
Reordering	1	4	6	3	3	2	2
Omission	1	3	3	3	5		1
Addition	10	6	16	9	7	5	4
Total	25	29	46	25	23	22	18

Fig. 12 Post-edited actions (PEA) of the "Apple Watch" user guide

Upon analyzing the results, it can be inferred that a significant proportion of the PEA instances recorded were attributed to alterations carried out by User 3, accounting for 46 modifications, whereas User 7 contributed to only 18 adjustments. The method of "Addition" was predominantly employed, while no alterations were detected within the prepositional classification. Depending on the class or subclass, certain sentences have undergone many alterations, including additions, reordering, and verb style changes. Because Lithuanian has so many different cases, it might be claimed that any modification made during post-editing causes the noun case to change. The majority of modifications to the Noun-Phrase category were implemented within the noun meaning choice subcategory. The observation was made that the precision of the machine output was inadequate due to the selective translation of certain button functions while others remained untranslated. Additionally, it was noted that the error originated from the source text, thereby rendering the output of the target machine inaccurate. The following are examples of post-editing actions for Noun-Phrase class:

Noun meaning choice

Ex 29 ST: [...] *Recovery rates* [...]
TT: [...] *Atsigavimo rodikliais* [...]
PE7 : [...] *Atsigavimo momentais* [...]

Ex 30 ST: [...] *Resting rate* [...]
TT: [...] *Poilsio dažnis* [...]
PE6 : [...] *Poilsio rodiklis* [...]

Ex 31 ST: [...] *Walking average rate* [...]
TT: [...] *Vidutinį ėjimo dažnį* [...]
PE2 : [...] *Vidutinį širdies ritmą ėjimo metu* [...]

Example 29 demonstrates that a modification was implemented by replacing the noun "rodikliais" with "momentais," which are distinct words with differing meanings. Examples 30 and 31 exhibit a similar phenomenon where nouns are substituted with other nouns that possess distinct semantic connotations. The other sub-class pertains to a case change and has undergone a total of 9 alterations. The examples are below:

Case change

Ex 32 ST: [...] *High or low heart rate* [...]
TT: [...] *Aukštą ar žemo širdies ritmo* [...]
PE4 : [...] *Aukštą ar žemą širdies ritmą* [...]

Ex 33 ST: [...] *See your heart rate* [...]
TT: [...] *Pamatykite savo širdies ritmą* [...]
PE3 : [...] *Stebėjimas* širdies *ritmo* [...]

The machine translation output in the given instance was deemed inaccurate due to the presence of two adjectives, namely "aukštą" and "žemo," which modify the same noun but are of distinct cases. All post-editors subsequently rectified this error. Example 33 also includes a change in case, but the primary reason for the modification was due to a change in the noun. The other Noun-Phrase sublass is adjective choice which was used 6 times. Comaparing to the previous sub-class it was more frequent. Below illustrated examples presents the modifications:

Adjective choice

Ex 34 ST: [...] *Current heart rate* [...]
TT: [...] *Esamą širdies ritmą* [...]
PE1 : [...] *Dabartinį širdies ritmą* [...]

Ex 35 ST: [...] *Current heart rate* [...]
TT: [...] *Esamą širdies ritmą* [...]
PE4 : [...] *Rodomą širdies ritmą* [...]

Both examples show the extract from the machine output but indicate the different post-editing choices. In the example 34 the TT word „esamą" was changed by post-editor 1 to „ dabartinį" and „ rodomą" by post-editor 4. The change in the example 34 seems to be more appropriate as it is closer to the ST phrase. One more of Noun-Phrase class is stylistic change of a noun which means that a synonym replaces noun with no change of meaning. In total there were 5 modifications performed and a couple of them are below:

Stylistic change of a noun

Ex 36 ST: [...] *The selected time period* [...]
TT: [...] *Pasirinktu laiku* [...]
PE2 : [...] *Pasirinktu laikotarpiu* [...]

Ex 37 ST: [...] *Walking average* [...]
TT: [...] *Vidutinio vaikščiojimo* [...]
PE6 : [...] *Vidutinio ėjimo* [...]
Ex 38 ST: [...] *To add Heart Rate to your summary* [...]
TT: [...] *Norėmi įtraukti “Heart Rate” į savo santrauką* [...]
PE7 : [...] *Norėdami įtraukti “Heart Rate“ į savo suvestinę* [...]

The alteration of a noun style is a technique that typically necessitates minimal post-editing attempts and is commonly executed to enhance the sentence or to be unit or phrase more appealing. The aforementioned examples demonstrate that all modifications are appropriate and convey the intended meaning equivalently. The final subclass experienced a total of 13 alterations, all of which involved multiple words. In comparison to the noun maening choice sub-class which underwent 15 modifications, it may be posited that this particular technique was similarly prevalent. Examples 39-40 indicated that units were changed by many other words. Several examples are below:

Multi-word change

Ex 39 ST: [...] *An important way to monitor how your body is doing* [...]
TT: [...] *Svarbus būdas stebėti kaip veikia jūsų kūnas* [...]
PE1 : [...] *Svarbus jūsų kūno būklės stebėjimas* [...]

Ex 40 ST: [...] *Recovery rates throughout the day;* [...]
TT: [...] *Atsigavimo rodiklius. per visą dieną;* [...]
PE7 : [...] *Atsigavimo rodiklius bet kuriuo metu* [...]

With respect to the Noun-Phrase category, it has been observed that the collective number of post-editing tasks carried out by all post-editors is 52. The noun denoting the act of selecting or making a decision was the most frequently employed, whereas the stylistic noun conveying alteration or modification was utilized merely on five occasions. The observed phenomenon suggests that the machine-generated translation output for the user guide necessitated significant modifications with respect to noun usage during the post-editing phase.

The other class is Verb-Phrase class and total of 37 modifications were made in relation to this particular class. The sub-class of verb agreement underwent 3 changes, while there were 6 variations in verb meaning choice and 28 modifications in verb stylistic usage. Upon comparing the “Addition” and “Omission” categories, which were not originally included in Blain’s et al. (2011) typology, it can be inferred that they were frequently utilized by the research participants. Undoubtedly, the addition strategy was more frequently employed. Below are the examples for Verb-Phrase and other classes, as well as examples containing errors in the form of incorrect case usage:

Verb meaning choice

Ex 41 ST: [...] *See your heart rat*e [...]
TT: [...] *Pamatykite savo širdies ritmą* [...]
PE1 : [...] *Pasitikrinkite savo širdies ritmą* [...]

Ex 42 ST: [...] *Workout View, then tap a workout* [...]
TT: [...] "*Treniruotės vaizdas", tada bakstelėkite treniruotę* [...]
PE5 : [...] *"Workout View", tada pasirinkite norimą treniruotę* [...]

Verb stylistic change

Ex 43 ST: [...] *Check your heart rate during a workout* [...]
TT: [...] *Patikrinkite savo širdies ritmą treniruotės metu* [...]
PE4 : [...] *Pasitikrinkite savo širdies ritmą treniruotės metu* [...]

Ex 44 ST: [...] *Tap My Watch, go to Workout* [...]
TT: [...] *Bakstelėkite "Mano laikrodis", eikite į "Treniruotė"* [...]
PE7 : [...] *Spauskite"Mano laikrodis", eikite į "Treniruotė* " [...]

Co-reference change

Ex 45 ST: [...] *Your can turn it back on* [...]
TT: [...] *Galite įjungti jį vėl įjungti* [...]
PE1-6 : [...] *Galite įjungti juos vėl* [...]

Omission

Ex 46 ST: [...] *throughout the day* [...]
TT: [...] *per visą dieną* [...]
PE5 : [...] *visą dieną* [...]

Ex 47 ST: [...] *Can turn it back on* [...]
TT: [...] *Galite įjungti jį vėl įjungti* [...]
PE3 : [...] *Galite vėl juos įjungti* [...]

Addition

Ex 48 ST: [...] *Tap Browse* [...]
TT: [...] *Bakstelėkite "Browse " (naršyti)* [...]
PE4 : [...] *Bakstelėkite "Browse " (liet. naršyti)* [...]

Ex 49 ST: [...] *Open the Apple Watch app on your iPhone* [...]
TT: [...] *Atidarykite "Apple Watch" programėlę iPhone* [...]
PE6 : [...] *Atidarykite "Apple Watch" programėlę iPhone telefone* [...]

Misc style-unnecessary changes

Ex 50 ST: [...] *Tap Browse at the bottom right* [...]
TT: [...] *Bakstelėkite “Browse” (naršyti) apačioje dešinėje* [...]
PE4 : [...] *Spūstelėkite dešinėje esančia nuorodą “Naršyt”i* [...]

Ex 51 ST: [...] *Tap Browse at the bottom right* [...]
TT: [...] *Bakstelėkite “Browse” (naršyti) apačioje dešinėje*[...]
PE6 : [...] *Bakstelėkite “Browse y” (angl. naršyti) apačioje dešinėje* [...]

MT output and port-editors’ errors

Ex 37 ST: [...] *If you’ve turned off heart rate data,* your *can turn* [...]
TT: [...] *Jei išjungėte širdies ritmo duomenis, galite įjungti jį vėl įjungti* [...]
PE3 : [...] *galite įjungti ir vėl išungti* [...]

Concerning the class of Verb-Phrase, significant modifications were observed in terms of stylistic variations in verbs. The prevalence of imperative verbs within the user guide text may be attributed to its intended purpose. As previously noted, there were errors present in the output of the machine translation. The majority of post-editors identified and rectified the co-reference error. The second scenario pertains to the button names, wherein some of them were translated into Lithuanian while others remained untranslated. Due to this factor, certain translators opted to translate all of them into the Lithuanian language, whereas others rectified the erroneous machine translation output and rendered all of the button functions into English, which was the appropriate choice of action. Upon comparing the “Addition” and “Omission” categories, which were not originally included in Blain et al.s’ (2011) typology, it can be inferred that they were frequently utilized by the research participants. Undoubtedly, the addition strategy was more frequently employed.

The findings of the analysis indicate that the post-editing of the user guide text was not a straightforward task, as evidenced by the total of 188 modifications made. This is in contrast to the initial text, which required 121 changes. The initial text underwent modifications primarily concerning the selection of noun phrases, whereas the latter text exhibited a greater number of alterations pertaining to the selection of verb phrases. Upon examination of the findings, it can be inferred that a significant proportion of the PEA events recorded are associated with alterations made to the “Addition” technique. No alterations were detected in the prepositional category within the second text. It could be posited that the machine-generated output of user guide text was deemed insufficient, thereby necessitating the inclusion of additional words by post-editors. An additional factor contributing to this phenomenon is the substantial dissimilarity between the English and Lithuanian languages. Additionally, it has been observed that certain modifications made by post-editors were either unnecessary or incorrect. For instance, in example 34, the post-editor displaced the abbreviation “liet” with “ang. naršyti”. In summary, it can be stated that certain participants performed the task proficiently, while others implemented modifications that were deemed unnecessary, resulting in increased cognitive effort.

Conclusions

In order to determine whether text content heterogeneity and complexity affect post-editing quality, this study presents a quantitative and qualitative comparative analysis of user activity data to identify temporal and cognitive efforts as well as the main issues encountered during post-editing of two different types of content texts. The most common DeepL translation program was used to translate two separate paragraphs from the Apple website, and the quantity of post-editing modifications and the quality of MT output were assessed using the Translog-II (automated tool for evaluating user data activity) and Blain et al's (2011) typology.

The following outcomes were obtained:

1) The current study offers a theoretical perspective on the established techniques and principles of post-editing machine translation, as well as the post-editing process and the characteristics of post-editors. Although recommendations and strategies may differ, it can be stated that post-editing is gaining greater prominence in the field of translation studies. This phenomenon can be attributed to an increasing acknowledgement of its importance in preserving the quality of machine translation results. The overview of the sources shows that the evaluation of post-editing endeavors can be accomplished through diverse methodologies and tools. Various commonly used methodologies utilized to measure post-editing endeavors include time monitoring, word tallies, error percentages, and quality evaluations. CAT tools such as SDL Trados, MemoQ, Wordfast, and Smartcat can be considered as feasible alternatives for evaluating post-editing efforts. Analyzing the theoretical sources, it is revealed that the intimate correlation between translation and post-editing mandates that proficient post-editors possess a varied range of competencies and expertise. Post-editing is a frequently utilized method in conjunction with machine translation to improve the efficiency and precision of the translation process.
2) Linear View, Pause Plot, and Statistics functions of the Translog-II software were employed for the purpose of conducting data analysis. The analysis of temporal effort was conducted through the utilization of the Statistics function, which provided insights into the duration of post-editing time. Additionally, cognitive effort was examined through the application of the Linear View function, which facilitated the examination of all user activity data. After conducting a comparison between the two texts, it was noted that the first text necessitated an extra four minutes for post-editing, even though the Translog-II software registered a lower number of events. The second text was found to have a mean execution time of 20 minutes, despite containing a higher quantity of documented occurrences. In summary, it is plausible that post-editors demonstrate a higher level of proficiency in translating instructional manuals. Another aspect to take into account is that further alterations were made as a result of unsatisfactory outcomes from machine translation. The results of the analysis suggest that the process of post-editing the user guide text was not a simple undertaking, as demonstrated by the 188 total modifications that were implemented. In contrast to the original text, a total of 121 modifications were necessary. The first text was modified mainly in terms of noun phrase selection, while the second text showed more changes related "Addition", "Reordering" strategies. Based on the analysis of the results, it can be deduced that a considerable percentage of the PEA recorded are linked to modifications implemented in the "Addition" methodology. One could argue that the output of user guide text generated by machines was

deemed inadequate, thus requiring post-editors to add supplementary words. One contributing factor to this phenomenon is the significant dissimilarity between the English and Lithuanian languages.

3) The primary problems encountered during the post-editing process of two distinct types of textual content relate to Button function names. Specifically, certain Button names were translated into Lithuanian, while others remained untranslated. As a result of this variable, some translators made the decision to translate all of the button functions into Lithuanian, while others corrected the inaccurate output generated by the machine translation and converted all of the Button functions into English, which was the suitable course of action. Furthermore, it can be inferred that the post-editing process of the "Apple" website research was more time-consuming due to the lengthy sentences, which evidently necessitated greater time and effort.
4) Upon comparing the two texts, it was observed that the initial text required an additional four minutes for post-editing, despite the Translog-II program recording fewer events. On average, the second text was executed within a duration of 20 minutes, yet encompassing a greater number of recorded events. In conclusion, it is possible that post-editors exhibit greater familiarity with the translation of user guides. Another factor to consider is that additional modifications were implemented due to inadequate machine translation results. The user guide underwent a total of 188 modifications. By comparison, the research article underwent a total of 121 modifications. In conclusion, the research article indicated a relatively lower frequency of modifications made to the noun class. The user guide has undergone further modifications pertaining to verb classification. A total of 57 modifications were executed through the utilization of the "Addition" technique in the user guide whereas the research article contained only 17. To conclude, "Addtition" strategy was more applied in user guide.

It is advisable for translators and post-editors to employ a diverse range of automated machine translation tools in order to determine the optimal quality of machine translation output. Moreover, it is recommended to adhere to all post-editing guidelines and strategies that are suggested by TAUS and language service providers. Post-editing is a key component of the translation process that can prove to be a formidable task. However, it can also be a stimulating and valuable experience that contributes to one's professional growth.

The present investigation is constrained by the extent of the post-edited text category, as it solely investigates the post-editing of a pair of articles. Thus, there exists a potential for further investigation to substitute the examination with diverse categories of source material, other language combinations, and automated translation programs.

List of References

1. Allen, J. (2003). Post-editing. *Benjamins Translation Library, 35*, 297-318 [viewed 25 February]. Retrieved from https://www.torrossa.com/en/resources/an/5001884#page=314
2. Alves, F., Koglin, A., &Mesa-Lao, (2016). Analysing the impact of interactive machine translation on post-editing effort. *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB,* , 77-94 [viewed 27 February]. Retrieved from https://link.springer.com/chapter/doi/10.1007/978-3-319-20358-4_4
3. Blain, F., Senellart, J., Schwenk, H., Plitt, M., & Roturier, J. (2011). Qualitative analysis of post-editing for high quality machine translation. In *Proceedings of Machine Translation Summit XIII* [viewed 28 February]. Retrieved from https://aclanthology.org/2011.mtsummit papers.17.pdf
4. Burchardt, A., Lommel, A., & Macketanz, V. (2021). A new deal for translation quality. *Universal Access in the Information Society, 20*(4), 701-715 [viewed 20 February]. Retrieved from https://link.springer.com/article/dois/10.1007/s10209-020-00736-510.1007/s10209-020-00736-5
5. Cambedda, G., Di Nunzio, G. M., & Nosilia, V. (2021). A study on automatic machine translation tools: A comparative error analysis between deepl and yandex for russian-italian medical translation. *Umanistica Digitale,* (10), 139-163. [viewed 21 February]. Retrieved from https://doi.org/10.6092/issn.2532-8816/12631
6. Carl, M. (2012). Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research. Paper presented at the *Lrec, 12* 4108-4112. [viewed 5 March]. Retrieved from http://www.lrec-conf.org/proceedings/lrec2012/pdf/614_Paper.pdf
7. Chauhan, S., & Daniel, P. (2022). A comprehensive survey on various fully automatic machine translation evaluation metrics. *Neural Processing Letters*, 1-55. [viewed 21 February]. Retrieved from https://link.springer.com/article/10.1007/s11063-022-10835-4
8. Chung-ling Shih. (2021). Re-Looking Into Machine Translation Errors and Post-Editing Strategies in a Changing High-Tech Context. *Compilation & Translation Review, 14*(2), 125-166. [viewed 25 March]. Retrieved from https://ctr.naer.edu.twv14.2/ctr140204.pdf
9. Colman, T., Fonteyne, M., Daems, J., & Macken, L. (2021). It’s all in the eyes: an eye tracking experiment to assess the readability of machine translated literature. In *31st Meeting of Computational Linguistics in the Netherlands (CLIN 31).* [viewed 4 March]. Retrieved from http://hdl.handle.net/1854/LU-8757156
10. Cui, Y., Liu, X., & Cheng, Y. (2023). A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment. *SAGE Open*, [viewed 20 February]. Retrieved from https://journals.sagepub.com/doi/full/10.1177/21582440231155849
11. Dabbadie, M., Hartley, A., &King, M. (2002). A hands-on study of the reliability and coherence of evaluation metrics. *Workshop at the LREC 2002 Conference, 8* [viewed 5 April]. Retrieved from https://mt-archive.net/00/LREC-2002-WS-MTEval.pdf#page=14

12. Dede, V. (2022). Temporal and Technical Effort in Post-editing Compared to Editing and Translation from Scratch. [viewed 5 February]. Retrieved from https://www.openaccess.hacettepe.edu.tr/xmlui/handle/11655/26391
13. Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the Second International Conference on Human Language Technology Research,* 138-145. [viewed 15 February]. Retrieved from https://dl.acm.org/doi/abs/10.5555/1289189.1289273
14. Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective. *The Journal of Specialised Translation, 11*(11), 52-74. [viewed 18 March]. Retrieved from https://jostrans.org/issue11/art_fiederer_obrien.pdf
15. Fomicheva, M. (2017). The Role of Human Reference Translation in Machine Translation Evaluation. [viewed 17 March]. Retrieved from https://www.tdx.cat/handle/10803/404987#page=1
16. Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics, 9,* 1460-1474. [viewed 5 February].Retrieved from https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00437/108866/Experts-Errors-and-Context-A-Large-Scale-Study-of
17. Garcia, I. (2012). A brief history of postediting and of research on postediting. *Revista Anglo Saxonica*, 291-310. [viewed 15 February]. Retrieved from https://researchdirect.westernsydney.edu.au/islandora/object/uws:13051/
18. Garvin, D. (1984). What Does "Product Quality" Really Mean? MIT Sloan Management Review.26.,25-43. [viewed15 February]. Retrieved from http://oqrm.org/English/What_does_product_quality_really_means.pdf
19. Graham, Y., Baldwin T., Alistair M., & Justin Z. (2013). Continuous measurement scales in human evaluation of machine translation. [viewed 11 February]. Retrieved from http://www.tara.tcd.ie/handle/2262/96112
20. Han, A. L., Wong, D. F., & Chao, L. S. (2012). LEPOR: A robust evaluation metric for machine translation with augmented factors. In *Proceedings of COLING 2012: Posters,* 441-450. [viewed 17 February]. Retrieved from https://aclanthology.org/C12-2044.pdf
21. Han, L., Jones, G. J., & Smeaton, A. F. (2021). Translation quality assessment: A brief survey on manual and automatic methods. *arXiv Preprint arXiv:2105.03311,* [viewed 15 February]. Retrieved from https://arxiv.org/abs/2105.03311
22. Herbig, N., Pal, S., van Genabith, J., & Krüger, A. (2019). Multi-modal approaches for post-editing machine translation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,*1-11. [viewed 18 February]. Retrieved from https://doi.org/10.1145/3290605.3300461
23. House, J. (1997). *Translation quality assessment: A model revisited.* Gunter Narr Verlag. [viewed 11 February]. Retrieved from https://books.google.lt/books?hl=lt&lr=&id=D16aYuTCBJ0C&oi=fnd&pg=PR7&dq=23.%09House,+J.+(1997).+Translation+quality+assessment:+A+model+revisited&ots=4YX_N24Nf7&sig=iVKUh0fQ3RWtUt4T9KbZnpy1BNE&redir_esc=y#v=onepage&q=23.%09House

%2C%20J.%20(1997).%20Translation%20quality%20assessment%3A%20A%20model%20revisited&f=false

24. Hu, K., & Cadwell, P. (2016). A comparative study of post-editing guidelines. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation,* 34206-34353. [viewed 16 February]. Retrieved from https://arxiv.org/abs/2105.03311
25. Hutchins, W. J. (1986). Machine translation: past, present, future. Citeseer. [viewed 11 February]. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9d415430967bff0f12bd1d68560e9afa46705c5b
26. ISO 18587:. (2017). *ISO 18587 Translation services — Post-editing of machine translation output — Requirements* [viewed 25 February]. Retrieved from https://www.iso.org/standard/62970.html
27. Kasperavičienė, R., Motiejūnienė, J., & Patašienė, I. (2020). Quality assessment of machine translation output. *Texto Livre, 13*(2), 271-285. [viewed 14 March]. Retrieved from https://www.redalyc.org/journal/5771/577164137008/577164137008.pdf
28. Kasperė, R., Motiejūnienė, J., Patašienė, I., Patašius, M., & Horbačauskienė, J. (2023). Is machine translation a dim technology for its users? An eye tracking study. *Frontiers in Psychology, 14,* 1-14. [viewed 15 February]. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9939441/
29. King, M., Popescu-Belis, A., & Hovy, E. (2003). FEMTI: creating and using a framework for MT evaluation. In *Proceedings of Machine Translation Summit IX: Papers,* [viewed 19 February]. Retrieved from https://aclanthology.org/2003.mtsummit-papers.30
30. Koby, G. S., Fields, P., Hague, D. R., Lommel, A., & Melby, A. (2014). Defining Landscape of Translation. *Tradumàtica,* (12), 413. 10.5565/rev/tradumatica.76 [viewed 20 January]. Retrieved from https://revistes.uab.cat/tradumatica/article/view/n12-melby-fields-hague-etal
31. Kocmi, T., Federmann, C. & Grundkiewicz, R. (2021). To ship or not to ship: An extensive evaluation of automatic metrics for machine translation. *arXiv Preprint arXiv:2107.10821,* [viewed 2 February]. Retrieved from https://doi.org/10.48550/arXiv.2107.10821
32. Koponen, M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the Seventh Workshop on Statistical Machine Translation,* 181-190. [viewed 15 February]. Retrieved from https://aclanthology.org/W12-3123.pdf
33. Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation, 25*(2) . [viewed 3 February]. Retrieved from https://www.jostrans.org/issue25/art_koponen.pdf
34. Krings, H. P. (2001). Repairing texts: Empirical investigations of machine translation post-editing processes. Kent State University Press. [viewed 15 February]. Retrieved from https://www.erudit.org/en/journals/meta/2002-v47-n3-meta693/008026ar/
35. Lacruz, I. (2017). The Handbook of Translation and Cognition. (pp. 386-401). Hoboken, NJ, USA. [viewed 9 February]. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781119241485?src=getftr

36. Läubli, S., & Germann, U. (2016). Statistical modelling and automatic tagging of human translation processes. *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB,* 155-181. [viewed 7 February]. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-20358-4_8
37. Lavie, A., & Agarwal, A. (2007). METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. *Wmt@ Acl, 2007*, 228-231. [viewed 15 February]. Retrieved from https://lt.upr.si/research_projects/mt_v_sloveniji/microsoft/ensl/eval/METEOR_en/meteor-0.6/papers/Lavie-Agarwal-2007-METEOR.pdf
38. Lin, C. (2004). Rouge: A package for automatic evaluation of summaries. In the *Text Summarization Branches Out,* 74-81. [viewed 26 February]. Retrieved from https://aclanthology.org/W04-1013.pdf
39. Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica,* (12), 455. [viewed 8 April]. Retrieved from https://revistes.uab.cat/tradumatica/article/view/n12-lommel-uzskoreit-burchardt
40. Marie, B. (2022). An Automatic Evaluation of the WMT22 General Machine Translation Task. *arXiv Preprint arXiv:2209.14172,* [viewed 4 April]. Retrieved from https://doi.org/10.48550/arXiv.2209.14172
41. Martínez, R. (2014). A deeper look into metrics for translation quality assessment (TQA): A case study. *Miscelánea: A Journal of English and American Studies, 49* [viewed 23 February]. Retrieved from https://www.miscelaneajournal.net/index.php/misc/article/view/170
42. Massardo I., Jaap M., & O'Brien S. (2016). *TAUS Post-Editing Guidelines*. The Netherlands: TAUS Signature Editions. [viewed 11 April]. Retrieved from https://www.taus.net/resources/reports/mt-post-editing-guidelines
43. Maučec, M. S., & Donaj, G. (2019). Machine translation and the evaluation of its quality. *Recent Trends in Computational Intelligence, 143* [viewed 4 April]. Retrieved from https://www.intechopen.com/chapters/68953
44. Moorkens, J., O'Brien, S. & Da Silva, I. A. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation, 29*, 267-284. [viewed 29 April]. Retrieved from https://link.springer.com/article/10.1007/s10590-015-9175-2
45. Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing (Volume 16)*. Language Science Press. [viewed 14 April]. Retrieved from https://library.oapen.org/handle/20.500.12657/52585
46. Novikova, J., Dušek, O., Curry, A. C., & Rieser, V. (2017). Why we need new evaluation metrics for NLG. *arXiv Preprint arXiv:1707.06875,* [viewed 4 April]. Retrieved from https://doi.org/10.18653/v1/D17-1237
47. O'Brien, S. (2002). Teaching post-editing: a proposal for course content. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation,* [viewed 4 April]. Retrieved from https://aclanthology.org/2002.eamt-1.11.pdf
48. O'Brien, S. (2011). Towards predicting post-editing productivity *Machine Translation, 25,* 197-215. [viewed 11 April]. Retrieved from https://link.springer.com/article/10.1007/s10590-011-9096-7

49. O'Brien, S. (2007). An empirical investigation of temporal and technical post-editing effort. *Translation and Interpreting Studies.the Journal of the American Translation and Interpreting Studies Association, 2*(1), 83-136. [viewed 17 April]. Retrieved from https://www.jbe-platform.com/content/journals/10.1075/tis.2.1.03ob
50. O'Brien, S., Balling, L. W., Carl, M., Simard, M., & Specia, L. (2014). Post-Editing of Machine Translation. Newcastle-upon-Tyne: Cambridge Scholars Publishing. [viewed 10 April]. Retrieved from https://books.google.lt/books?hl=lt&lr=&id=W3IxBwAAQBAJ&oi=fnd&pg=PR5&dq=50.%09O%27Brien,+S.,+Balling,+L.+W.,+Carl,+M.,+Simard,+M.,+%26+Specia,+L.+(2014).+Post-Editing+of+Machine+Translation.+&ots=f9mqOzrGBi&sig=WxWKu0sjf5w0DZol8OxFvDXqRKo&redir_esc=y#v=onepage&q&f=false
51. O'Brien, S., & Simard, M. (2014). Introduction to special issue on post-editing. *Machine Translation*, *28*(3), 159-164. [viewed 6 April]. Retrieved from https://link.springer.com/article/10.1007/s10590-014-9166-8
52. Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics,* 311-318. 8 [viewed 2 April]. Retrieved from https://aclanthology.org/P02-1040.pdf
53. Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation,* 392-395. [viewed 3 March]. Retrieved from https://aclanthology.org/W15-3049.pdf
54. Popović, M. (2020). Informative manual evaluation of machine translation output. In *Proceedings of the 28th International Conference on Computational Linguistics,* 5059-5069. [viewed 13 March]. Retrieved from https://aclanthology.org/2014.eamt-1.41.pdf
55. Popović, M., Lommel, A., Burchardt, A., Avramidis, E., & Uszkoreit, H. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, 191-198. [viewed 9 March]. Retrieved from https://aclanthology.org/2014.eamt-1.41.pdf
56. Popović, M., & Ney, H. (2011). Towards Automatic Error Analysis of Machine Translation Output. *Computational Linguistics - Association for Computational Linguistics, 37*(4), 657-688. [viewed 13 April]. Retrieved from https://doi.org/10.1162/COLI_a_00072
57. Povilaitienė, M., & Kasperė, R. (2022). Machine translation for post-editing practices. *Науковий Часопис Національного Педагогічного Університету Імені М.П.Драгоманова.Серія 9.Сучасні Тенденції Розвитку Мов,* (24), 47-62. [viewed 29 April]. Retrieved from https://doi.org/10.31392/NPU-nc.series9.2022.24.04
58. Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. *arXiv Preprint arXiv:2009.09025,* [viewed 27 April]. Retrieved from https://doi.org/10.48550/arXiv.2009.09025
59. Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu’s concepts. *The Journal of Specialised Translation, 31,* 201-216. [viewed 26 January]. Retrieved from https://pure.port.ac.uk/ws/files/13162370/art_sakamoto.pdf

60. Sellam, T., Das, D., & Parikh, A. P. (2020). BLEURT: Learning robust metrics for text generation. *arXiv Preprint arXiv:2004.04696*, [viewed 3 April]. Retrieved from https://doi.org/10.48550/arXiv.2004.04696
61. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, 223-231. [viewed 25 April]. Retrieved from https://aclanthology.org/2006.amta-papers.25
62. Specia, L., Hajlaoui, N., Hallett, C., & Aziz, W. (2011). Predicting machine translation adequacy. In *Proceedings of Machine Translation Summit XIII: Papers,* [viewed 5 April]. Retrieved from https://aclanthology.org/2011.mtsummit-papers.58.pdf
63. Stanchev, P., Wang, W., & Ney, H. (2019). EED: Extended edit distance measure for machine translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1),* 514-520. [viewed 7 March]. Retrieved from https://aclanthology.org/W19-5359
64. Vieira, L. N. (2019). Post-editing of machine translation. *The Routledge handbook of translation and technology* (pp. 319-336). Routledge. [viewed 8 April]. Retrieved from https://www.taylorfrancis.com/chapters/edit/10.4324/9781315311258-22/post-editing-machine-translation-lucas-nunes-vieira
65. Walker, C., & Federici, F. M. (2018). Eye Tracking and Multidisciplinary Studies on Translation. The Netherlands: John Benjamins Publishing Company. [viewed 18 April]. Retrieved from https://www.torrossa.com/en/resources/an/5002367
66. Wang, W., Peter, J., Rosendahl, H., & Ney, H. (2016). Character: Translation edit rate on character level. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers,* 505-510. [viewed 25 April]. Retrieved from https://aclanthology.org/W16-2342.pdf
67. Wisniewski, G., Singh, A. K., Segal, N., & François, Y. (2013). Design and analysis of a large corpus of post-edited translations: quality estimation, failure analysis and the variability of post-edition. In *Machine Translation Summit XIV,* [viewed 16 April]. Retrieved from https://hal.science/hal-03748614/
68. Zaretskaya, A. (2017). Machine Translation Post-Editing at TransPerfect–the 'Human'Side of the Process. *Revista Tradumàtica: Tecnologies De La Traducció,* (15), 116-123. [viewed 25 April]. Retrieved from https://doi.org/10.5565/rev/tradumatica.201
69. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. *arXiv Preprint arXiv:1904.09675*, [viewed 29 Apri]. Retrieved from https://arxiv.org/abs/1904.09675

Appendices

Appendix 1. Survey on post-editing experience in machine translation

1. Your educational background in translation:

- Bachelor's degree
- Master's degree
- Other (please specify)

2. Your translation experience:

- Less than 1 year
- Between 1 and 3 years
- Between 3 and 5 years
- More than 5 years

3. Do you use machine translation tools in professional translation?

- Yes
- No (if you answered No, your survey is closed, thank you for your answers)

4. How often do you use machine translation tools in your translations?

- Always
- Often
- Rarely
- Very rarely

5. Which machine translation tools do you use most often?

- Google Translate
- DeepL
- Microsoft translator
- Tilde
- CAT tools
- Other (please specify)

6. Do you do post-editing of machine translations?

- Yes
- No

7. If yes, what is your post-editing experience?

- Less than 1 year
- Between 1 and 3 years
- Between 3 and 5 years
- More than 5 years

8. What are the most common problems and difficulties you encounter in post-editing machine translations?

Thank you for your answers!

Appendix 2. Apklausa apie mašininio vertimo postredagavimo patirtį

1. Jūsų išsilavinimas vertimo srityje:

- Bakalauro laipsnis
- Magistro laipsnis
- Kita (nurodykite)

2. Jūsų vertimo patirtis:

- Mažiau kaip 1 metai
- Nuo 1 iki 3 metų
- Nuo 3 iki 5 metų
- Daugiau kaip 5 metai

3. Ar naudojate mašininio vertimo įrankius profesionaliame vertime?

- Taip
- Ne (jei atsakėte Ne , jūsų apklausa baigta, dėkojame už atsakymus)

4. Kaip dažnai jūsų vertimuose naudojate mašininio vertimo įrankius?

- Visada
- Dažnai
- Retai
- Labai retai

5. Kokius mašininio vertimo įrankius dažniausiai naudojate?

- Google Translate
- DeepL
- Microsoft translator
- Tilde
- CAT įrankius
- Kita (nurodykite)

6. Ar atliekate mašininio vertimo postredagavimą?

- Taip
- Ne

7. Jei atsakėte “Taip “, kokia yra jūsų postredagavimo patirtis?

- Mažiau kaip 1 metai

- Nuo 1 iki 3 metų
- Nuo 3 iki 5 metų
- Daugiau kaip 5 metai

8. Su kokiomis problemomis ir sunkumais dažniausiai susiduriate atlikdami mašininio vertimo postredagavimą?

Dėkojame už atsakymus!

Appendix 3. Source text of “Apple “ website research article

Retrieved from: https://www.apple.com/newsroom/2023/02/with-apple-watch-researchers-explore-new-frontiers-in-heart-health/

With Apple Watch, researchers explore new frontiers in heart

In a single day, the heart of an average, healthy adult beats more than 100,000 times. Beat by beat, day by day, a picture begins to emerge — an image that goes largely unseen. Apple Watch can help make the invisible, visible. With heart health features — including high and low heart notifications, Cardio Fitness, irregular rhythm notifications, the ECG app, and AFib History — Apple gives users an ever-developing view of their health with actionable insights.

The same advanced technology that provides insights to help users better understand their health also has the potential to open the door to discovery for the research and medical communities. Since Apple launched ResearchKit and CareKit in 2015, researchers, clinicians, and developers have found innovative new ways to study, track, and treat a broad range of conditions.

To further drive discoveries that improve health at scale, Apple launched the Investigator Support Program. Through this program, Apple provides researchers with Apple Watch devices, enabling them to break new ground in health research, including the scientific understanding of the heart.

On Apple Watch Series 4 or later, the ECG app can record heartbeat and rhythm using the electrical heart sensor. The ECG waveform stored in the Health app can be shared as a PDF.

Associate professor Rachel Conyers and Dr. Claudia Toro are senior pediatric oncologists from Melbourne, Australia, primarily spending their days caring for children in a tertiary pediatric oncology clinic and researching toxicities related to children’s cancer therapies within the Murdoch Children’s Research Institute. Together they are looking at how treatment can impact heart rhythm and are trying to identify innovative ways to intervene. The inspiration for their work comes from their patients — both success stories and heartbreak.

Appendix 4. MT output of “Apple” website research article

Retrieved from: https://www.apple.com/newsroom/2023/02/with-apple-watch-researchers-explore-new-frontiers-in-heart-health/

Su "Apple Watch" mokslininkai tyrinėja naujas širdies sveikatos ribas

Per vieną dieną vidutinio sveiko suaugusio žmogaus širdis plaka daugiau nei 100 000 kartų. Diena po dūžio, diena po dienos, ima ryškėti vaizdas - vaizdas, kuris dažniausiai lieka nepastebėtas. Laikrodis "Apple Watch" gali padėti nematomą padaryti matomu. Su širdies sveikatos funkcijomis, įskaitant pranešimus apie aukštą ir žemą širdies ritmą, "Cardio Fitness", pranešimus apie nereguliarų ritmą, EKG programą ir "AFib History", "Apple" suteikia naudotojams nuolat tobulėjantį jų sveikatos vaizdą su įžvalgomis, kurias galima panaudoti.

Ta pati pažangi technologija, kuri teikia įžvalgas, padedančias naudotojams geriau suprasti savo sveikatą, taip pat gali atverti duris atradimams mokslinių tyrimų ir medicinos bendruomenėms. Nuo tada, kai 2015 m. "Apple" pristatė "ResearchKit" ir "CareKit", tyrėjai, gydytojai ir kūrėjai rado inovatyvių naujų būdų tirti, stebėti ir gydyti įvairias būkles.

Siekdama toliau skatinti atradimus, kurie gerina sveikatą plačiu mastu, "Apple" pradėjo Tyrėjų paramos programą. Pagal šią programą "Apple" suteikia tyrėjams "Apple Watch" įrenginius, leidžiančius jiems atverti naujus sveikatos tyrimų, įskaitant mokslinį širdies supratimą, kelius.

Laikrodžiuose "Apple Watch Series 4" ar vėlesniuose įrenginiuose EKG programa galima fiksuoti širdies plakimą ir ritmą naudojant elektrinį širdies jutiklį. Programėlėje "Health" išsaugotą EKG bangos formą galima bendrinti PDF formatu.

Docentė Rachel Conyers ir daktarė Claudia Toro yra vyresniosios vaikų onkologės iš Melburno (Australija), daugiausia dienų praleidžiančios slaugydamos vaikus tretinio lygio vaikų onkologijos klinikoje ir tyrinėdamos toksiškumą, susijusį su vaikų vėžio gydymo metodais Murdocho vaikų tyrimų institute. Kartu jie stebi, kaip gydymas gali paveikti širdies ritmą, ir bando rasti naujoviškų intervencijos būdų. Įkvėpimo jų darbui semiasi iš savo pacientų - tiek sėkmės istorijų, tiek širdgėlos.

Appendix 5. Source text of "Apple Watch" user guide (p.292)

Retrieved from: https://help.apple.com/pdf/watch/8/en_US/apple-watch-user-guide-watchos8.pdf

Check your heart rate on Apple Watch (p.169-170)

Your heart rate is an important way to monitor how your body is doing. You can check your heart rate during a workout; see your resting, walking, breathe, workout, and recovery rates throughout the day; or take a new reading at any time.

See your heart rate

Open the Heart Rate app on your Apple Watch to view your current heart rate, resting

rate, and walking average rate. Your Apple Watch continues measuring your heart rate as long as you're wearing it.

Check your heart rate during a workout

By default, your current heart rate shows on the Multiple Metric workout view. To customize

which metrics appear during a workout, follow these steps:

1. Open the Apple Watch app on your iPhone.

2. Tap My Watch, go to Workout > Workout View, then tap a workout.

For more information, see Start a workout on Apple Watch.

See a graph of your heart rate data

1. Open the Health app on your iPhone.

2. Tap Browse at the bottom right, tap Heart, then tap Heart Rate.

3. To add Heart Rate to your Summary, swipe up, then tap Add to Favorites.

You can see your heart rate over the last hour, day, week, month, or year. Tap Show More

Heart Rate Data and you can also see the range of your heart rate during the selected time

period; your resting, walking average, workout, sleep, and Breathe rates; and any high or low heart rate notifications.

Turn on heart rate data

By default, your Apple Watch monitors your heart rate for the Heart Rate app, workouts, and Breath and Reflect sessions. If you’ve turned off heart rate data, your can turn it back on.

Appendix 6. MT output of Apple Watch user guide (p.292)

Retrieved from: https://help.apple.com/pdf/watch/8/en_US/apple-watch-user-guide-watchos8.pdf

Širdies ritmo tikrinimas "Apple Watch

Širdies ritmas yra svarbus būdas stebėti, kaip veikia jūsų kūnas. Galite patikrinti savo širdies ritmą treniruotės metu; matyti savo ramybės, ėjimo, kvėpavimo, treniruotės ir atsigavimo rodiklius per visą dieną; arba bet kuriuo metu nuskaityti naujus rodmenis.

Pamatykite savo širdies ritmą

Atidarykite "Apple Watch" programėlę "Heart Rate" ir peržiūrėkite esamą širdies ritmą, poilsio

dažnį ir vidutinį ėjimo dažnį. Laikrodis "Apple Watch" matuoja jūsų širdies ritmą tol, kol jį nešiojate.

Patikrinkite savo širdies ritmą treniruotės metu

Pagal numatytuosius nustatymus jūsų dabartinis širdies ritmas rodomas "Multiple Metric" treniruotės rodinyje. Jei norite pritaikyti kuriuos rodiklius rodyti treniruotės metu, atlikite šiuos veiksmus:

1. Atidarykite "Apple Watch" programėlę iPhone.

2. Bakstelėkite "Mano laikrodis", eikite į "Treniruotė" > "Treniruotės vaizdas", tada bakstelėkite treniruotę.

Daugiau informacijos rasite skyriuje "Treniruotės pradžia "Apple Watch" laikrodyje".

Širdies ritmo duomenų grafiko peržiūra

1. Atidarykite "Health" programėlę iPhone.

2. Bakstelėkite "Browse" (naršyti) apačioje dešinėje, bakstelėkite "Heart" (širdis), tada bakstelėkite "Heart Rate" (širdies ritmas).

3. Norėdami įtraukti "Heart Rate" į savo santrauką, vilkite aukštyn, tada bakstelėkite "Add to Favorites".

Galite matyti savo širdies ritmą per pastarąją valandą, dieną, savaitę, mėnesį ar metus. Bakstelėkite Rodyti daugiau Širdies ritmo duomenys ir taip pat galite pamatyti savo širdies ritmo diapazoną pasirinktu laiku. laikotarpį; savo ramybės, vidutinį vaikščiojimo, treniruotės, miego ir kvėpavimo dažnį; taip pat bet kokį aukštą ar žemo širdies ritmo pranešimus.

Širdies ritmo duomenų įjungimas

Pagal numatytuosius nustatymus "Apple Watch" stebi jūsų širdies ritmą programėlei "Heart Rate", treniruotes, ir "Kvėpavimo" bei "Atspindžių" seansus. Jei širdies ritmo duomenis išjungėte, galite įjungti jį vėl įjungti.