



Kaunas University of Technology
Faculty of Social Sciences, Arts and Humanities

Expert and Non-Expert Human Acceptability of Machine Translation from English to Lithuanian

Master's Final Degree Project

Greta Pinkevičiūtė

Project author

Prof. dr. Ramunė Kasperė

Supervisor

Kaunas, 2021



Kaunas University of Technology
Faculty of Social Sciences, Arts and Humanities

Expert and Non-Expert Human Acceptability of Machine Translation from English to Lithuanian

Master's Final Degree Project
Translation and Localisation of Technical Texts (6211NX031)

Greta Pinkevičiūtė

Project author

Prof. dr. Ramunė Kasperė

Supervisor

Prof. dr. Saulius Keturakis

Reviewer

Kaunas, 2021



Kaunas University of Technology
Faculty of Social Sciences, Arts and Humanities
Greta Pinkevičiūtė

Expert and Non-Expert Human Acceptability of Machine Translation from English to Lithuanian

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.

Greta Pinkevičiūtė

Confirmed electronically

Pinkevičiūtė, Greta. Expert and Non-Expert Human Acceptability of Machine Translation from English to Lithuanian. 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: acceptability, human evaluation, machine translation, translation studies, translation quality assessment

Kaunas, 2021. 57 p.

Summary

This paper presents a research on human acceptability of machine translation from English to Lithuanian. The novelty of this research is that it provides an insight on professionals as well as basic users' (experts and non-experts') acceptability of a random machine translation system's output for the English–Lithuanian language pair. The relevance of the research is that it analyses the aspects that make a translation acceptable and acknowledges the issues that have the most influence on acceptability as perceived by machine translation users. Therefore, the object of the research is human acceptability of machine translation from English to Lithuanian. The aim is to analyse acceptability of machine translation from English to Lithuanian and its influential factors as assessed by machine translation users. The set-out objectives to meet the aim are as follow:

1. to discuss human acceptability in translation studies;
2. to discuss translation quality assessment metrics;
3. to analyse experts' and non-experts' acceptability of machine translation;
4. to indicate the most problematic machine translation aspects by assessing the machine translation quality;
5. to evaluate the factors that have the most impact on the machine translation quality and acceptability as assessed by its users.

The research includes the descriptive, comparative, qualitative and quantitative statistical analysis methods. The main tool to analyse acceptability of machine translation is the acceptability model proposed by Castilho (2016) who proposed that acceptability is measured by considering three elements: usability, quality and satisfaction. Moreover, Multidimensional Quality Metrics (MQM) introduced by Lommel, Burchardt and Uszkoreit (2015b) are employed to assess machine translation quality by considering accuracy, fluency, style and locale convention dimensions.

Acceptability is assessed via two different surveys by 20 experts and 70 non-experts. In this research, experts are individuals who have experience in translation and/or degree in translation or languages, whilst non-experts are basic machine translation users whose background is unknown.

The results of the analysis show that machine translation from English to Lithuanian is deemed unacceptable by all surveys' respondents. Even though non-experts share more positive views towards satisfaction and usability of machine translation, they agree with experts that the low quality of a translation is the main factor that makes it unacceptable. Further analysis also shows that machine translation users are mostly affected by critical accuracy errors present in a translation, whereas fluency issues are identified as disturbing yet not major, and issues of style and locale convention are not perceived as having any influence over the comprehensibility of machine translation by both

experts and non-experts. Taking everything into account, machine translation from English to Lithuanian is deemed unacceptable by its users, the main reason of which is low quality of machine translation. The low quality and incomprehensibility of machine translation is mostly influenced by critical accuracy errors as perceived by both experts and non-experts.

The research is comprised of two parts. The theoretical part of the work discusses theoretical aspects of human acceptability in general, human acceptability of machine translation, and methods to measure acceptability. It also reviews translation quality assessment methods. The analytical part of the work presents an analysis of experts and non-experts' acceptability of machine translation from English to Lithuanian and the factors that have an impact on it as perceived by its users with a focus on the quality element.

Pinkevičiūtė, Greta. Mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtumas ekspertų ir ne ekspertų vertinimu. Magistro studijų baigiamasis projektas / vadovė prof. dr. Ramunė Kasperė; Kauno technologijos universitetas, Socialinių, humanitarinių mokslų ir menų fakultetas.

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

Reikšminiai žodžiai: priimtumas, žmogaus vertinimas, mašininiais vertimas, vertimo studijos, vertimo kokybės vertinimas

Kaunas, 2021. 57 p.

Santrauka

Šiame darbe tiriamas mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtumas. Temos naujumas susijęs su mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtumu vertėjų ir įprastų naudotojų (ekspertų ir ne ekspertų) vertinimu. Ši tema aktuali todėl, nes yra analizuojami vertimo priimtimumui įtakos turintys veiksniai ir apžvelgiamos problemos, kurios lemia mašininio vertimo priimtimumą jo naudotojo atžvilgiu. Taigi, šio tyrimo objektas yra mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtumas. Tikslas yra išanalizuoti mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtimumą ir jam įtaką darančius veiksnius naudotojų vertinimu. Nusistatyti darbo uždaviniai yra:

1. aptarti vertimo priimtimumo aspektus;
2. aptarti vertimo kokybės vertinimo metodus;
3. ištirti mašininio vertimo priimtimumą ekspertų ir ne ekspertų vertinimu;
4. išskirti problematiškiausius mašininio vertimo aspektus įvertinus mašininio vertimo kokybę;
5. išanalizuoti veiksnius, turinčius daugiausiai įtakos mašininio vertimo priimtimumui vertimo naudotojų atžvilgiu.

Tyrime pasitelkiami aprašomasis, lyginamasis, kokybinis ir kiekybinis statistinis analizės metodai. Pagrindinė mašininio vertimo priimtimumo tyrimo priemonė yra tyrėjos Castilho (2016) priimtimumo modelis, kuriuo teigiama, jog priimtumas matuojamas vertinant vertimo naudojamumą, kokybę ir patitenkinimą vertimu. Kokybė yra vertinama pagal Lommel, Burchardt ir Uszkoreit (2015b) pristatyta *Multidimensional Quality Metrics* (MQM) metodiką, analizuojant tikslumo, sklandumo, stiliaus ir pritaikymo lokalei dimensijas.

Priimtumas vertinamas 20-ies ekspertų ir 70-ies ne ekspertų pasitelkus apklausas. Šiame tyrime, ekspertai yra asmenys, turintys vertimo patirties, baigę vertimo ar kalbų studijas, o ne ekspertai yra įprasti mašininio vertimo sistemų naudotojai, kurių patirtys bei pareigos nėra žinomi.

Tyrimo rezultatai rodo, kad mašininis vertimas iš anglų kalbos į lietuvių kalbą nėra priimtinas apklausų respondentų nuomone. Nors ne ekspertai geriau vertina naudojamumą ir reiškia didesnę patitenkinimą mašininio vertimu nei ekspertai, visi naudotojai sutinka, kad žema vertimo kokybė yra pagrindinis veiksnys, darantis vertimą nepriimtiniu. Rezultatai taip pat rodo, kad mašininio vertimo naudotojų priimtimumui didžiausią įtaką turi kritinės tikslumo klaidos, kai tuo tarpu sklandumo klaidos yra trikdančios, tačiau ne itin svarios, o stiliaus ir pritaikomumo lokalei klaidos yra vertinamos kaip neturinčios įtakos mašininio vertimo kokybei ir jo priimtimumui ekspertų ir ne ekspertų vertinimu. Taigi galima teigti, jog mašininis vertimas iš anglų kalbos į lietuvių kalbą yra nepriimtinas jo naudotojų atžvilgiu, o tam didžiausią įtaką daro žema vertimo kokybė. Ekspertų ir ne

ekspertų vertinimų, ši žema mašininio vertimo kokybė ir vertimo nesuprantamumas kyla dėl kritinių tikslumo klaidų.

Tyrimą sudaro dvi dalys. Teorinėje darbo dalyje aptariame teoriniai priimtimumo aspektai, mašininio vertimo priimtimumas ir priimtimumo vertinimo metodai. Šioje dalyje taip pat aptariami vertimo kokybės vertinimo metodai. Analitinėje darbo dalyje pateikiama mašininio vertimo iš anglų kalbos į lietuvių kalbą priimtimumo ekspertų ir ne ekspertų vertinimu ir faktorių turinčių įtakos priimtimumui analizė atsižvelgiant į kokybės elementą.

Table of contents

List of figures	9
List of tables	10
Introduction	11
1. Theoretical aspects of machine translation acceptability and translation quality assessment	13
1.1. Human acceptability and its measurement	13
1.1.1. Acceptability in translation studies	13
1.1.2. Methods to measure machine translation acceptability	18
1.2. Machine translation errors and quality assessment	21
1.2.1. Machine translation quality assessment methods	21
1.2.2. Standardised translation quality assessment systems	23
2. Human acceptability of machine translation	28
2.1. Methodology	28
2.2. Acceptability of machine translation as assessed by experts and non-experts	31
2.2.1. Expert and non-expert satisfaction with the machine translation	32
2.2.2. Expert and non-expert opinion on the machine translation usability	33
2.2.3. Expert and non-expert opinion on the machine translation quality	35
2.3. Analysis of machine translation errors	36
2.4. Influence of produced errors on machine translation acceptability	42
2.4.1. Relationship between errors and experts' acceptability of machine translation	42
2.4.2. Relationship between errors and non-experts' acceptability of machine translation	46
Conclusions	54
List of references	56
List of information sources	60
Appendices	61
Appendix 1. Source text and target text	61
Appendix 2. Experts' survey on acceptability	63
Appendix 3. Non-experts' survey on acceptability	70
Appendix 4. Machine translation quality assessment by employing Multidimensional Quality Metrics	76

List of figures

Fig. 1. Acceptability model (Castilho, 2016).....	16
Fig. 2. Dimensions of Multidimensional Quality Metrics (Lommel et al., 2015a).....	24
Fig. 3. Hierarchical structure of Multidimensional Quality Metrics dimension of Accuracy (Lommel et al., 2015a).....	25
Fig. 4. Answers to statement 1: The main idea of the text is clear.....	32
Fig. 5. Answers to statement 2: The text is written in an orderly, understandable way	33
Fig. 6. Answers to statement 3: The text is useful, informative.....	34
Fig. 7. Answers to statement 4: I would trust the source of the text	34
Fig. 8. Answers to statement 5: The text sounds natural	35
Fig. 9. Answers to statement 6: The text is correct, and its quality is excellent	35
Fig. 10. Number of issues in the analysed machine translation	36
Fig. 11. Issues of the Accuracy dimension	37
Fig. 12. Issues of the Fluency dimension.....	38
Fig. 13. Issues of the Locale convention dimension.....	40

List of tables

Table 1. One-way MANOVA results on machine translation acceptability	31
Table 2. Experts' evaluation of machine translation.....	43
Table 3. Non-experts' evaluation of machine translation	47

Introduction

The occurrence of the automated translation systems has presented a very fast way to obtain information presented in a foreign language, and the availability of machine translation systems has made them an inseparable element of daily lives. The ease of use, accessibility and speed have popularised the systems to an extent that the services provided by professionals have started to lose their priority. However, the preference and popularity of machine translation systems do not imply that the output they provide is of the highest quality possible and that it would not cause any difficulties to its target reader, i.e., that it would not be disturbed in any way that would have negative influence on users' acceptability of a translation. Acceptability has become an important and widely discussed metric when assessing translation in a way that it concerns not only the linguistic transfer of ideas present in a source text to a target text, but also target text users' subjective opinion about machine translation output. The subjectivity of the matter makes the use of this measure a complex task as what is acceptable to one individual, might be completely unacceptable to the other because of the needs or expectations that different machine translation systems' users have. For this reason, it is important to understand the aspects that make a translation acceptable and to acknowledge the issues that have the most influence on the acceptability of machine translation.

Even though acceptability of machine translation has been the object of many discussions for some time now, considerably few studies were carried out to analyse acceptability of machine translation in Lithuanian as a target language. Therefore, the **novelty** of this research is that it provides an insight on professionals as well as basic user's acceptability of a random machine translation system's output for the English–Lithuanian language pair. It allows to draw conclusions on whether the satisfaction, usability and quality elements are met and to indicate the most problematic factors that influence the overall acceptability of the machine translation output. The **relevance** of the research is that it analyses the aspects that make a translation acceptable and acknowledges the issues that have the most influence on acceptability as perceived by machine translation users.

Thus, the **problem** raised in this paper is how acceptable is machine translation output from English to Lithuanian and what factors influence the acceptability.

The **object** of the analysis is human acceptability of machine translation from English to Lithuanian.

Therefore, **the aim** is to analyse acceptability of machine translation from English to Lithuanian and its influential factors as assessed by machine translation users.

The set-out **objectives** to meet the aim are as follow:

1. to discuss human acceptability in translation studies;
2. to discuss translation quality assessment metrics;
3. to analyse experts' and non-experts' acceptability of machine translation;
4. to indicate the most problematic machine translation aspects by assessing the machine translation quality;
5. to evaluate the factors that have the most impact on the machine translation quality and acceptability as assessed by its users.

The following hypotheses were raised:

1. Machine translation from English to Lithuanian is deemed more acceptable by non-experts than experts;
2. Quality issues impact acceptability of machine translation the most.

The research includes the descriptive, comparative, qualitative and quantitative statistical analysis methods that help to assess the acceptability of the machine translation from English to Lithuanian and the factors that have impact on it.

During the master studies, the results of the individually carried-out research *Eye-Tracking Experiments in the Evaluation of the Machine- and Human-Produced Translation Quality* were presented at the students' scientific conference *SMILES 2020: SOCIAL SCIENCES, ARTS AND HUMANITIES IN CONTEMPORARY SOCIETY*.

1. Theoretical aspects of machine translation acceptability and translation quality assessment

The extensive usage of machine translation systems has occurred due to the systems' compatibility with the fast rhythm of modern life, in which low costs and great speed are of the greatest importance. The orientation towards fast results may suggest that quality has become a less important aspect which has lost its significance and does not have a considerable influence on the acceptability of translation and that the users of machine translation systems' output view machine translation as acceptable despite the flaws that are present in it. For this reason, it is important to find out the elements that make a translation acceptable as viewed by its users and to discuss the aspects that make machine translation less acceptable. These aspects are analysed in detail in the following sections of the theoretical part of the work.

1.1. Human acceptability and its measurement

The extensive usage of modern technologies is decreasing the need for human services relentlessly; this tendency arises from the fact that technologies create a product or provide a service much faster and at significantly lower costs than humans do (Way, 2018, p. 160). Although it may signal that human intelligence is losing its importance in certain spheres and ways, it must not be forgotten that only human professionals are able to improve the technologies that have become an inherent part of everyday life. One of the spheres that require special attention from human specialists is translation. Humans are able to connect and communicate with people from all across the globe; we can also reach information written in any language possible. Nonetheless, this easy access to foreign content as well as the possibility to communicate with people from different parts of the world is constrained by the language barrier that can be partly reduced or completely eliminated by the help of human or machine translation. Since people have become oriented towards a quick result at the lowest price possible, machine translation systems have become a preferred method to obtain the information provided in a foreign language. For this reason, questions arise whether a usually highly desired quality of produced translations has become a less important aspect, or the notion of quality as perceived by users of translations has been altered; should a translation be of an outstanding quality to be accepted by its readers. It is, therefore, important to know and understand what characteristics a translation must have so that it is acceptable to its target user, and to find out the ways of how acceptability can be measured.

1.1.1. Acceptability in translation studies

A great variety of the acceptability notion definitions has caused a different understanding on the meaning of it. First and foremost, acceptability is defined as the fourth of the seven standards of textuality all of which transform a text into a "communicate occurrence" (De Beaugrande, & Dressler, 1981, p. 6). The fourth standard – acceptability – is defined as "the text receiver's attitude that the set of occurrences should constitute a cohesive and coherent text having some use or relevance for the receiver" (De Beaugrande, & Dressler, 1981, p. 9). This "attitude" automatically signals that acceptability is a subjective measure that is assessed by a target user of a text. However, it also indicates that the standard of acceptability consists of specific elements that are taken into account when deeming a text acceptable or unacceptable.

A great contribution to the study of acceptability has been made by Chomsky (1965), who has defined the notion as a "perfectly natural and immediately comprehensible [utterances] without paper-and-pencil analysis, and in no way bizarre or outlandish" (p. 9). According to the scholar, a utterance is

perceived as acceptable if it contains no unnatural phrasing as well as unknown words that would have any negative influence on the comprehension of the idea presented in it. As well as De Beaugrande and Dressler (1981), Chomsky (1965) defines acceptability as a general feature of any type of text. Chomsky (1965) also draws attention to the fact that acceptability is often compared to grammaticality; however, grammaticality is “only one of many factors that interact to determine acceptability” (p. 10). The author emphasises that the main difference between the two notions is that the concept of acceptability appertains in the “study of performance”, and grammaticality appertains in the “study of competence” (Chomsky, 1965, p. 9), and while grammaticality has a certain degree of influence to acceptability, the opposite is quite impossible as its definition would indicate that there can be unacceptable grammatical sentences. This would lead to an exhausting analysis and even reconsideration of standard grammar rules. Moreover, Chomsky (1965) proposes that grammaticality is a far more important notion than acceptability, and while the latter can be measured by applying various tests and experiments, it is quite impossible that a “sufficient operational criterion might be invented” to analyse and measure grammaticality (p. 10). For this reason, it is important to realise the difference between the two notions before carrying out analysis on either of them as the results of a study which employs the two notions interchangeably would provide valuable results neither on the grammaticality, nor on the acceptability of a subject in question.

The term of “acceptability” in translation studies was used by Van Slype (1979). Unlike De Beaugrande and Dressler (1981) and Chomsky (1965) who defined acceptability as a general feature of any type of text (see above), Van Slype (1979) adjusts the definition to the translation field by describing acceptability as “a subjective assessment of the extent to which a translation is acceptable to its final user” (p. 92). The subjectivity of this cognitive level assessment suggests that measuring acceptability is a tough and time-consuming task that requires additional resources to be used so that the results are as accurate as possible. Moreover, the researcher suggests that a translation is deemed acceptable if it meets the needs of its target user; the acceptability is also dependent on the purpose as well as a text type of the source text (Van Slype, 1979, p. 35). Van Slype (1979) also points out that measuring machine translation acceptability by its users is advantageous since the translation’s quality is judged by its target reader and that the criteria of acceptability are simple: a translation is either acceptable or not acceptable, and that it “relates to the actual purpose of the operation” (p. 112), whereas the main disadvantage of measurement of acceptability of machine translation is that in order to obtain accurate and reliable results, a big number of documents as well as its users (or experiment participants) is required (Van Slype, 1979, p. 113). Mentioning this scholar’s input into the study of acceptability is noteworthy as many of researchers (e.g., Neubert & Shreve, 1992; Roturier, 2006; Castilho, Doherty, Gaspari, & Moorkens, 2018; Kasperavičienė, Motiejūnienė, & Patašienė, 2020, etc.) continue to employ and share the valuable insights of Van Slype’s (1979) in their research even to this day.

Acceptability is also understood as a translation’s adherence to certain norms and needs. Neubert and Shreve (1992) propose that even though there are no universal norms of acceptability that apply to different text types of different purposes, texts should “possess particular textual features, including standard grammatical and lexical patterning” to be accepted by its readers (Neubert & Shreve, 1992, p. 73). It is also discussed that a translation can only be viewed as acceptable if it is rendered with regards to target text standards as well as expectations which indicates that a translator must be aware of the target language and its peculiarities in order to produce an acceptable translation (Neubert & Shreve, 1992, p. 73). If textual standards are not regarded, the receiver of the text will not be able to

determine the purpose of it; therefore, the produced translation will not be regarded as a text and its linguistic communication will not be achieved (Neubert & Shreve, 1992, p. 73). All in all, the scholars propose that a text is acceptable if it meets standards – therefore, needs – of its target user which coincides with the previously presented idea of acceptability proposed by Van Slype (1976).

Acceptability is also discussed to be highly dependent on the issues present in a translation. Williams (2009) discusses that there is not yet a definition proposed that would describe what is acceptable and what is not (p. 3); however, the same author in his earlier work *Translation Quality Assessment: an argumentation-centred approach* proposes the notion of acceptability threshold, i.e., “the level of tolerance of errors” (Williams, 2004, p. XVIII). This indicates that a text or a utterance can be perceived as acceptable and the expectations or needs of its users can be met if the number of errors and the severity of errors a text or utterance contains does not disrupt the comprehension of a presented idea; therefore, “an acceptable translation is one that fully conveys the argument macrostructure of the source text and is therefore free of critical defects” (Williams, 2009, p. 14). This understanding of acceptability is similar to the idea proposed by PACTE (2009) research group, who view acceptability as quality of a translation product (p. 216). The scholars draw attention to the accurateness of information present in a translation as compared with its source text (PACTE, 2008, p. 20). More specifically, the scholars analyse acceptability in terms of meaning of the source text, function of the translation, and language use (PACTE, 2008, p. 20). Taking into account the three criteria, translation is then deemed:

- acceptable – a translation meets all three criteria,
- semi-acceptable – a translation meets some of the criteria or meets the criteria to some extent,
- not acceptable – a translation meets none of the three criteria (PACTE, 2008, p. 20).

When describing acceptability, PACTE (2008, 2009) and Williams (2009) draw attention not only to a product, but also the importance of a source text; therefore, a translation can only be deemed acceptable if it fully and qualitatively conveys the information present in a source text. This notion of the quality of a rendered translation also coincides with the idea proposed by Toury (1995). The scholar states that a translator should be aware of the standards and rules (similarly as stated by the previously mentioned Neubert and Shreve (1992) of both source and target language as those have direct influence on the manner a translation is produced, i.e., a product may either be “adequate” if it is to be based on rules of the source language and culture or “acceptable” if it is to be based on rules of the target language and culture (Castillo, 2015, p. 73). While this explanation may resemble foreignization and domestication (Venuti, 2008, p. 20), this emphasises the importance of how information should be conveyed and what has to be sacrificed so that a translation is accepted by its target readers.

Acceptability has also been analysed from the point of view of style (Lassen, 2003). The scholar defines acceptability as an interactive process between a reader and the text, in which the reader “assesses whether the text has qualities which make the style appropriate for the situation” (Lassen, 2003, p. xv). This suggests that acceptability is directly concerned with the style of a text; an aspect which has not been noted by many researchers. Moreover, Lassen (2003) proposes that acceptability concerns both grammatical and stylistic acceptability; therefore, acceptability as a whole is perceived as the target user’s opinion on the produced text that has been modified in terms of cohesiveness and coherency to become usable (p. xv). The definition that is used by the scholar coincides with the

definition of the seventh standard of textuality as proposed by De Beaugrande and Dressler (1981) and also suggests one of the possible elements of acceptability – i.e., usability.

In his work, Roturier (2006) borrows the previously mentioned definition provided by De Beaugrande and Dressler (1981) to refer to acceptability as “the text receiver’s attitude that the set of occurrences should constitute a cohesive and coherent text having some use or relevance for the receiver” (p. 7) (see p. 10) adding that acceptability does not only refer to the relevance of a text, but also the “manner in which its textual characteristics are going to be accepted, tolerated, or rejected by its receivers” (Roturier, 2006, p. 4). Therefore, a machine-produced translation may still be viewed as acceptable even if it contains errors; however, only a few disturbances are tolerated (Roturier, 2006, p. 157). This idea also correlates with the suggestion proposed by the previously analysed research of Williams (2003; 2009).

In presence of a number of different definitions and vague characteristics of the notion of acceptability, it, nonetheless, can be discussed in terms of different elements that it constitutes. Castilho (2016) has proposed a different definition of acceptability in translation by emphasising three of its substituents: usability, quality and satisfaction (p. xviii). Castilho (2016) suggests that acceptability is addressed through the following research question: “What factors influence acceptability levels of a machine translated text for the end user?” (p. 3), noting that acceptability is not a simple, single notion that is to be analysed as a unit; it comprises of different substituents that have to be evaluated separately. The scholar emphasises the importance of the source text when evaluating a target text as acceptability of the former may have a significant influence on the usability of the latter (Castilho, 2016, p. 15), an idea which is similar to one proposed by PACTE (2008) (see p. 12).

The proposed acceptability model that Castilho (2016) has used in her research is shown in Fig. 1 that is presented below.



Fig. 1. Acceptability model (Castilho, 2016)

The scholar has drawn up the acceptability model based on the ideas proposed by Nielsen (1993) in which acceptability is composed of various categories including usability, De Beaugrande and Dessler’s (1981) idea of acceptability as a relevance of the text (p. 6) as well as Roturier’s (2006) suggestion that acceptability refers to the extent to which “textual characteristics [of a text] are going to be accepted, tolerated and rejected by its receiver” (p. 4). The author systemises the proposed acceptability ideas into the three already mentioned acceptability criteria: usability, quality and satisfaction (Castilho, 2016, p. 58).

The term of usability in this acceptability model (see Fig. 1) is understood as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” (International Organization for Standardization [ISO], 2002,

p. 2). The definition is complemented by Byrne (2006), who notes that usability refers to the readability and comprehension of a text as well as the ease to perform a task that is required by the text (p. 201). Nonetheless, in her acceptability model Castilho (2016) mainly bases usability definition on the one provided by Soujanen, Koskinen and Tuominen (2015), who point out a text can be deemed usable “if users can typically use it in a satisfactory manner in the context in which it was intended” (p. 14), moving away from the functional definition that usability is usually described as. Therefore, it could be said that usability of a translation reflects its user’s willingness to use the text to meet their needs in whichever way – either to obtain some information from it, or to use the text to fulfil a specific task.

The criterion of quality in Castilho (2016) acceptability model (see Fig. 1) refers to the extent to which errors present in the translation do not have a negative influence on comprehensibility and the extent to which a translation can be considered as “good enough” (p. 61). As defined by the scholar, the main factors that have an impact on the quality criterion are adequacy, fluency, syntax and grammar, and style (Castilho, 2016, p. 90). These factors show that quality depends on how well a translation is rendered linguistically.

To define the third criterion of accessibility, Castilho (2016) refers to the definition of satisfaction that is proposed by International Organization for Standardization (ISO, 1998), i.e., “freedom from discomfort, and positive attitudes towards the use of the product” (p. 2). The scholar also notes that satisfaction is usually measured in human-computer interaction field; however, the definition can be adjusted to fit the needs of research in translation field by viewing usability not how “pleasant it is to use the system” (Nielsen, 1993, p. 33), but how pleasant it is to use, i.e., to read a product of a translation process (Castilho, 2016, p. 61). Moreover, the scholar combines the definition proposed by ISO (1998) with the definition of suggested by Rubin and Chisnell (2011) who propose that satisfaction is “user’s perceptions, feelings, and opinions of the product, usually captured through both written and oral questioning” (p. 4). It indicates that the criterion of satisfaction is a very subjective component of acceptability; however, this does not imply that it is by any means less important additive that should be overlooked when evaluating acceptability of a translation.

Therefore, the proposed acceptability model (see Fig. 1) suggests that a translation can be deemed acceptable by its users only if a) it is easy to understand as well as perform tasks indicated in it, b) it contains no significant errors or the number of errors is not preventing an end user to understand the information provided in it, and c) it is pleasant to read as well to use. For the purposes of the following research, this model will be used to measure acceptability as it clearly defines the elements of the subjective concept.

Taking everything into account, a translation is acceptable if it has no significant errors, causes no comprehension issues to its target reader, and qualitatively conveys the information present in a source text. Notwithstanding the proposed adequately clear acceptability criteria consisting of usability, quality and satisfaction, measuring acceptability can prove to be a difficult task. Measuring the acceptability of a translation is by no means free of subjectivity as this aspect can only be evaluated by people; subjectivity arises from the fact that what can be acceptable to one person, might be completely unacceptable for another. For this reason, it is important to discuss the methods that are used to measure acceptability; this is further analysed in the following section.

1.1.2. Methods to measure machine translation acceptability

The previous section has shown that the definition of acceptability varies; however, it is agreed that acceptability is mainly a user's opinion on a product's usability, quality and satisfaction. The different definitions as well as the subjectivity of the matter make evaluation of acceptability a difficult task as researchers choose to measure it with different methods. Therefore, it is important to overview and compare the different ways of how results on human acceptability of machine translation are achieved.

Following Van Slype's (1979) definition of acceptability as a "subjective assessment" (p. 92), the scholar proposes to measure acceptability by online surveys (Van Slype, 1979, p. 13). Even though the scholar does not provide any further explanation on the survey's questions criteria, measuring acceptability via online surveys has become a popular way to evaluate if an end user finds a translation acceptable. These and other methods are discussed below.

One of the attempts to measure acceptability has been done by Coughlin (2003) who has aimed to investigate the correlation between human and automated metrics to evaluate the quality of the machine translation output. The scholar employs a scale from 1 to 4 to assess acceptability of translations, i.e.:

1. Ideal – grammatically correct translation with all information transferred from source text.
2. Acceptable – stylistically or grammatically unusual, yet comprehensible and with accurate transfer of information.
3. Possibly acceptable – possibly comprehensible; with some transfer of information.
4. Unacceptable – completely incomprehensible, with little or no transfer of information (Coughlin, 2003, p. 84).

It must be noted that the scholar does not define the term of acceptability in her research; however, the statements of the presented scale show that acceptability is understood in terms of grammar, style and comprehensibility and that it is evaluated by a text's target users. The factors influencing the acceptability of translation coincides with the factors that impact the criterion of quality in Castilho (2016) acceptability model (see. Fig. 1). Therefore, the scale proposed by Coughlin (2003) suggests that quality is the main factor that influences the comprehensibility, thus acceptability of a translation and that quality and comprehensibility completely depend on each other.

Acceptability is also found to be evaluated by providing offline surveys that help to obtain translation's users opinion on a presented machine translation. The previously mentioned Lassen (2003) (see p. 13) employed an offline survey to evaluate both accessibility and acceptability of technical documentation. The research has shown that the term of acceptability may often mean different things to different respondents; more specifically, acceptability "may imply grammaticality to some respondents, while it may imply stylistic acceptability to others" (Lassen, 2003, p. 81). This draws attention to the fact that acceptability must be clearly defined before asking respondents to evaluate it in order to reduce a risk of significant deviations as well as misunderstanding of a task.

Roturier (2006) has also followed the idea proposed by Van Slype (1979) to measure acceptability by a survey of final users. The researcher proposes a four-value scale to evaluate the output of machine translation, i.e.:

1. Excellent MT output (E) – output is satisfactory, syntactically correct, it uses proper terminology. The output conveys source text’s information accurately.
2. Good MT output (G) – output contains minor grammatical mistakes that do not cause comprehensibility issues.
3. Medium MT output (M) – output contains significant errors that cause comprehensibility issues.
4. Poor MT output (P) – output is incomprehensible (Roturier, 2006, p. 86).

It is proposed that a translation can often be evaluated as “Excellent” or “Good” not only because it contains no significant errors, but also because of end users’ “in-depth knowledge about the topic” (Roturier, 2006, p. 87). This implies that there is a need to choose both the participants of the experiment as well as the text as a grammatically correct sentence can still “be difficult to read for several reasons: unfamiliar vocabulary, complex syntactic structure, syntactic or semantic ambiguity, etc.” (Stymne et al., 2012, p. 1083). Therefore, what could seem to be a flawless sentence corresponding to all grammar rules, can still cause difficulties due to insufficient knowledge of the given text language or topic. This idea was also discussed by Zamanian and Heydari (2012), who defined this phenomenon as the “reader-text mismatch” (p. 43). Therefore, to receive accurate results as well as to avoid possible significant deviations, it is important that a relevant text is chosen and that “the evaluation of documents is performed by genuine users” (Roturier, 2006, p. 149) who express interest in the topic at least to some extent. Furthermore, the respondents were presented with question such as “Was this document useful?” with only two possible answer options, i.e., “Yes” and “No” (Roturier, 2006, p. 158) as to avoid ambiguity as well as loss of interest. According to researchers, such options such as “I don’t know” can often be used by respondents who are not interested in the topic or who become tired of the survey, whereas open questions can often be unanswered (Galesic, 2006, p. 325). It is, therefore, important not only to choose suitable statements on acceptability, but also to provide only those answer options that would not have significantly negative impact on the overall results of the study.

A different method to measure translation acceptability is by providing a post-task questionnaire to participants of an experiment. Doherty and O’Brien (2014), Gondra (2018), Kasperavičienė et al. (2020), Castilho (2016) in their research have used a 5-point Likert scale¹ to obtain respondents’ insight on acceptability of a translation (it must be noted that Castilho and O’Brien (2016) have used the Likert scale only to assess satisfaction of a machine translation product (p. 312). For instance, the Likert scale was used in Castilho (2016) research who presented statements (e.g., “The instructions were comprehensible”, “I was satisfied with the instructions provided”, “The instructions were usable”, etc.) to respondents in a post-task survey to obtain experiment participants’ satisfaction with machine translation (p. 81). Moreover, the Likert scale was also used to assess translation quality in terms of adequacy, fluency, syntax, and other aspects as perceived by participants (Castilho, 2016, p. 103). Furthermore, Kasperavičienė et al. (2020) also employed the Likert scale to get insight on the respondents’ opinion on the presented text’s acceptability in terms of usability, quality and satisfaction (p. 275), i.e., the scholars have employed the acceptability model proposed by Castilho (2016) (see Fig. 1) by providing statements to respondents to assess the three elements of acceptability (e.g., statement “The main idea of translated text was easy to understand” to measure satisfaction, statement “The translation is suitable for publication.” to measure usability, statement “The sentences in the text sound natural.” to measure quality (Kasperavičienė et al., 2020, p. 285). It must also be

¹ 1 – disagree, 2 – somewhat disagree, 3 – neither disagree, nor agree, 4 – somewhat agree, 5 – agree.

mentioned that the research carried out by Kasperavičienė et al. (2020) is one of the two research studies (other one is the research of Daubarienė and Ziezytė (2013) in which acceptability of machine translation output for English–Lithuanian language direction was, to some extent, analysed. While the study performed by Daubarienė and Ziezytė (2013) has shown that machine translation output does not comply with standards of textuality and is not deemed acceptable by its target users (p. 60), firm conclusions on acceptability of machine translation in Lithuanian as a target language cannot be made due to scarce number of the carried-out research on the given topic. Therefore, further studies should be performed to get more insight.

A different yet widely used method to assess acceptability of translations is magnitude estimation (ME). A great contribution to the analysis of this method has been done by Weskott and Fanselow (2011) who compared ME with binary and seven-point judgements to find out whether ME method provides more valid and well-reasoned results on acceptability (p. 250). Magnitude estimation has originated in psychophysics but has been moved to the field of syntax by Bard, Robertson, and Sorace (1996) (Weskott & Fanselow, 2011, p. 249). This unusual move from one field to another has made ME a very different cognitive task in which an experiment's participants are asked to assess acceptability of a sentence "by using the acceptability of a different sentence as a unit of measure" (Sprouse, 2011, p. 274), i.e., participants need to state which sentence is better, and they also need to indicate how many times a sentence is better (Featherston, 2005, p. 1528). Even though the ME method has been successfully used in some research (e.g., Bader & Häussler, 2010; Keller, 2000; Sprouse & Almeida, 2012; Kertz, 2013 and other), Weskott and Fanselow (2011) found the hypothesis that "ME judgments are more informative than seven point judgments" (p. 269) is not true, as the results of the three carried-out experiments on word order variation in German did not show any particular differences that would have significant influence on the overall results of acceptability measuring. Therefore, the comparative analysis done by Weskott and Fanselow (2011) shows that applying the complex magnitude estimation method does not provide any additional information that would prove to be more useful than the usual measure systems, e.g., binary or seven-point (or 5-point) measures. Since there are no significant differences between the methods, it is, therefore, more recommended to use the simpler acceptability judgment methods purely for the sake of the participants' interest (see more information on p. 16).

The presented research show that acceptability can be measured via online surveys by presenting its participants with statements that they should evaluate according to a chosen scale (e.g., Likert scale, binary measures (i.e., "yes/no" or "acceptable/unacceptable"). Another possible acceptability measuring method is magnitude estimation that allows drawing conclusion on acceptability by asking the experiment's participants to evaluate as well as compare the presented sentences. However, it is found that the magnitude estimation method does not provide more informative results; therefore, using simpler methods such as the Likert scale or binary measures is recommended for their simplicity as well as applicability to online surveys or post-task questionnaires since those are recommended and are the most popular way to obtain experiments' participants acceptability judgements on presented translations. Also, as stated by Reips (2002), online experiments are advantageous because of the ease to access a large number of participants, highly voluntarily participation, avoidance of time constraints and high external validity (p. 245). These advantages characteristics of online surveys make them a suitable method to measure acceptability; however, due to the subjectivity of the matter and possible deviations that should be eliminated, a large number of respondents is needed to obtain general and valid results.

The analysed research shows that no single method has been introduced and used by researchers who assess acceptability due to the subjectivity of the notion as well as the lack of a single definition. However, it is also found that acceptability is mainly understood as target users' opinion on the text that is best obtained by questioning the users directly i.e., by providing questionnaires on acceptability of a text. A great variety of different methods to measure acceptability (e.g., binary measures, 5-point, 7-point scales, magnitude estimation) are employed; nonetheless, it has been found that the acceptability of a machine translation can be very well measured by providing the simplest forms of questionnaires – e.g., binary measures – as they provide the same results as that of a more complex magnitude estimation method which requires a relatively long time to be completed. Moreover, it is apparent that there are aspects that have to be considered before measuring acceptability, e.g., the eligibility of experiment participants and the number of them, the relevance of the topic of a text, the compatibility of a text and its reader, and the straightforwardness of presented questions. Considering these aspects before carrying out an acceptability experiment may prove to deliver more valid results. Most importantly, all of the discussed research carried out on acceptability show that the errors present in a translation always have an impact on the overall acceptability. For this reason, attention should be drawn to the types of mistakes and the weights that they have on acceptability of a translation that can also indicate the weakest aspects of modern translation systems that are the most noticeable by its users and that should be noted by content creators; this aspect is further analysed in the next section.

1.2. Machine translation errors and quality assessment

As it was discussed in the previous sections, acceptability is, above all, directly linked to quality, i.e., a text is deemed acceptable only if its quality meets the needs of its reader. Also, as it is pointed out by Williams (2003) and Roturier (2006), a translation can be viewed as acceptable if the types of errors or number of errors that it contains do not have significant impact on the understanding of a text (see p. 12). Therefore, to have a better understanding on how the element of quality may impact the acceptability of a translation, it is important to find out what types of errors cause the most trouble for a translation's target reader. On the other hand, assessing translation quality is an especially difficult task since “there is no single objective way to measure quality” (Drugan, 2013, p. 35); however, the complexity of this issue does not reduce the desire to create a method that would help to assess translation quality as objectively as possible (Koby, Fields, Hague, Lommel, & Melby, 2014, p. 416). These attempts to systemise machine translation errors and to move forward a universal translation quality assessment (hereinafter referred to as TQA) system are discussed in the further sections.

1.2.1. Machine translation quality assessment methods

There have been attempts made to systemise and propose error classification taxonomies that could be employed when analysing translation quality. For instance, Vilar, Xu, D'Haro, and Ney (2006) have distinguished five base categories of machine translation errors, those being missing words, word order, incorrect words, unknown words, and punctuation (p. 698). This taxonomy has been successfully applied in various research (e.g., Bojar, Ercegovčević, Popel and Zaidan, 2011; Farrús, Costa-Jussà, Mariño and Fonollosa, 2012; Temnikova, 2010; Temnikova, Zghouani, Vogel and Habash, 2016; Kasperavičienė et al., 2020, etc.) analysing machine translation.

A very significant research has been carried out by Temnikova (2010) and Temnikova et al. (2016) who redesigned the taxonomy proposed by Vilar et al. (2006) according to the cognitive effort that is required to post-edit the machine-produced errors. The research has found out that:

- morphological errors (i.e., incorrect word form) have the least impact on the processing of translation;
- lexical errors (i.e., incorrect synonym, incorrect word, missing word, idiomatic expression) require medium cognitive effort to correct them;
- syntactic errors (i.e., wrong or missing punctuation, word order) require the most cognitive effort in post-editing as they influence the understanding of the whole text the most (Temnikova, 2010, p. 3488).

This finding shares similarities with the results of the study carried out by Federico, Negri, Bentivogli, and Turchi (2014) who found that different error types have a different impact on human quality scores. One of the results of the study concludes that lexical and syntactic errors have the most influence on the overall quality (Federico et al., 2014, p. 1649); however, it is noted that the frequency of an error type does not necessarily become an indicator of a low-quality score as perceived by the text's users (Popović, 2018, p. 134). Thus, it becomes apparent that different error types have a different impact on a translation's quality as well as the cognitive effort needed in post-editing stage in order to render an acceptable translation.

Another different yet noteworthy research was carried out by Williams (2009) who has drawn up a TQA model with the idea that inaccurate transfer of an argument present in a source text makes a translation unusable as the mission of the texts is altered (p. 13). The scholar, therefore, distinguishes three categories of errors:

1. Critical – errors that misrepresent the argument of the source text and make a translation unusable.
2. Major – transfer errors that have an impact on the understandability of a translation but does not deem translation unusable.
3. Minor – other errors (Williams, 2009, p. 13).

The presented error classification acts as a basis for acceptability of a translation (see also section 1.1.1), claiming that “an acceptable translation is one that fully conveys the argument macrostructure of the source text and is therefore free of critical defects” (Williams, 2009, p. 14). This means that a translation can be deemed acceptable even though there are errors in it; however, the errors should not be too significant.

Nonetheless, the big variety of already-introduced noteworthy error classification taxonomies does not eliminate the subjectivity of translation quality assessment procedure. The big number of different taxonomies allows researchers to variate between them and apply different quality assessment methods in their studies. As a result, if one machine translation output is to be assessed by employing several different methods as well as taxonomies, the results may vary. For this reason, it is important to discuss that there is a need of a universal TQA system that would allow researchers to study the errors present in the analysed translation output in regard to their type and the influence that they have on the overall quality of a translation.

1.2.2. Standardised translation quality assessment systems

To eliminate the subjectivity of quality evaluation, a demand for standardised translation quality assessment systems has arisen, and those have been started to be developed. One of the first standards – SAE J2450 – was introduced by Society of Automotive Engineers (SAE) with the aim to establish “a standard quality metric for the ... automotive service information (Sirena, 2004). This standard presented 6 error types (i.e., wrong term, misspelling, omission, word structure or agreement, syntactic, punctuation and miscellaneous errors) with two severity levels (major / minor). The creation and implementation of this standard was followed by the introduction of LISA QA Model that was one of the most popular choices in TQA until its closure in 2011 (Snow, 2015, p. 73) and that has been further applied in various research even after that (Castilho et al., p. 15). This standard included from 18 to 21 categories (e.g., mistranslation, consistency as well as layout, graphics errors and many more) (the discrepancy arises from its documentation and interface (Lommel, 2018, p. 112) and had three severity levels (minor / major / critical); errors and their weights then lead “to an overall score for the whole translation task” (Castilho et al., 2018, p. 15) which determine a translation status as pass or fail. LISA QA Model still remains one of the most used models to assess the quality of translation; however, its users modify and adapt the categorisation of errors categories as they are often more appropriate for different text types or scenarios (Lommel, 2018, p. 112).

Other two noteworthy TQA standards that are introduced and widely used are Dynamic Quality Framework (DQF) and Multidimensional Quality Metrics (MQM). DQF Error Typology was developed by the Translation Automation User Society (TAUS), and it consists of six error types: accuracy, linguistic, terminology, style, country standards, and layout. It also has four other categories that are used to appoint issues that are not translation errors: query implementation, client edit, repeat, and kudos (Lommel, 2018, p. 124). This typology also contains four severity levels (critical, major and minor and neutral) to which numbers are assigned (Lommel, 2018, p. 124). The severity levels allow researchers to obtain a general translation quality score and to draw conclusions on which errors cause the most comprehension issues.

Furthermore, the mentioned MQM typology was developed in EU funded QTLaunchPad project to identify and eliminate the drawback of the previous translation quality assessment systems (Lommel, Burchardt & Uszkoreit, 2015, p. 458). It was created after closely comparing the briefly discussed and many more other standards and systems that have been introduced to move towards a universal TQA system (Lommel, 2018, p. 114). Its aim is “to provide a set of criteria which can be used to assess the quality of translations” (Lommel et al., 2015b). MQM Typology is tailored towards a functionalist approach that suggests that only those parts of the complex method can be used to meet the purpose of a research (Uszkoreit & Lommel, 2013; Castilho et al., 2018, p. 16). It also must be noted that the MQM can be used to evaluate not only machine translation output, but also human translations as well as a source text because of the metrics’ versatility (Castilho et al., 2018, p. 17). This universality of the metrics makes it a desirable and preferential tool in the analysis of machine translation quality, and it is seen to be gaining significant popularity in recent research (e.g., Castilho et al., 2017; Ortiz-Boix & Matamala, 2017; O’Brien, Balling, Carl, Simard & Specia, 2014; Klubička, Toral, & Sánchez-Cartagena, 2017; Klubička, Toral, & Sánchez-Cartagena, 2018; Vardaro, Schaeffer, & Hansen-Schirra, 2019; Ye & Toral, 2020; Freitag et al., 2021, etc.).

The MQM Typology consists of 182 issue types (Lommel, Burchardt, Görög, Uszkoreit & Melby, 2015a) that are a part of a very detailed hierarchy. The very top level of the hierarchy consists of eight dimensions that are figuratively shown in Fig. 2.

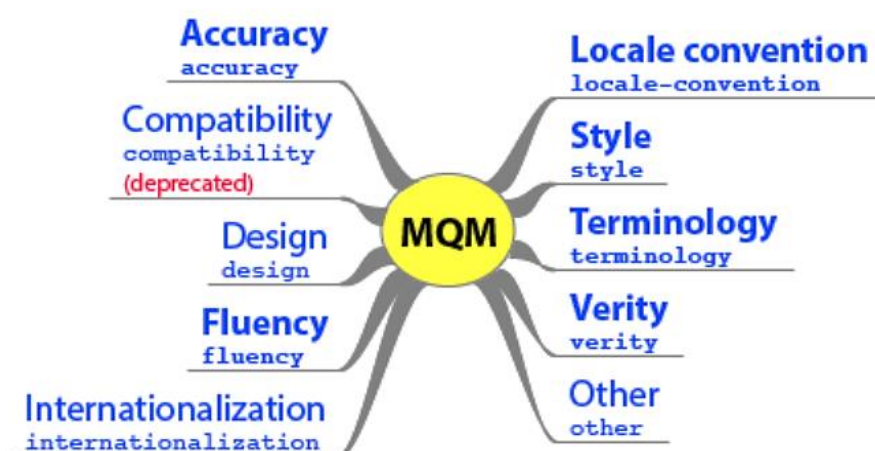


Fig. 2. Dimensions of Multidimensional Quality Metrics (Lommel et al., 2015a)

These dimensions are defined as follows:

- **Accuracy** (18 issues). Issues of meaning compatibility between source and target content. Its main categories² are addition, omission, untranslated, mistranslation.
- **Compatibility** (deprecated). Issues related to “legacy metrics that are not considered appropriate for general use in MQM” (Lommel et. al., 2015a).
- **Design** (33 issues). Issues related to the presentation of a text, i.e., desktop publishing. A few of its categories are hyphenation, overall design, local formatting, etc.
- **Fluency** (39 issues). Issues related to linguistic completeness of a text, despite if its translation or not. A few of its categories are grammar, spelling, cohesion, inconsistency, etc.
- **Internationalisation** (49 issues). Issues related to source text preparation for translation and localisation. A few of its categories are language-specific tool support, user interface internationalisation, graphical aspects, etc.
- **Locale convention** (14 issues). Issues related to the formal compliance of a translation to with locale-specific conventions, e.g., a format of addresses or dates. A few of its categories are calendar type, name format, time format, currency format, etc.
- **Style** (7 issues). Issues related to adherence to style guides as well as the informal “feel” of a text. A few of its categories are register, company style, inconsistent style, third party style, etc.
- **Terminology** (7 issues). Issues related to adherence to specified terminology. Its main categories are inconsistency with termbase, inconsistency with domain, and inconsistent use of terminology.
- **Verity** (7 issues). Issues related to the content suitability to the target audience. A few of its categories are completeness, end-user suitability, legal requirements, etc. (Lommel, 2018, p. 118; Lommel et al., 2015a).

² full list of issues of Accuracy dimension as well as other dimensions that are explained below is available on Lommel, Burchardt & Uszkoreit (2015a).

To have a better understanding on the MQM categorisation of issues, an attention must be drawn to the strict hierarchy of its dimensions. A hierarchy of Accuracy dimension is illustrated in Fig. 3.

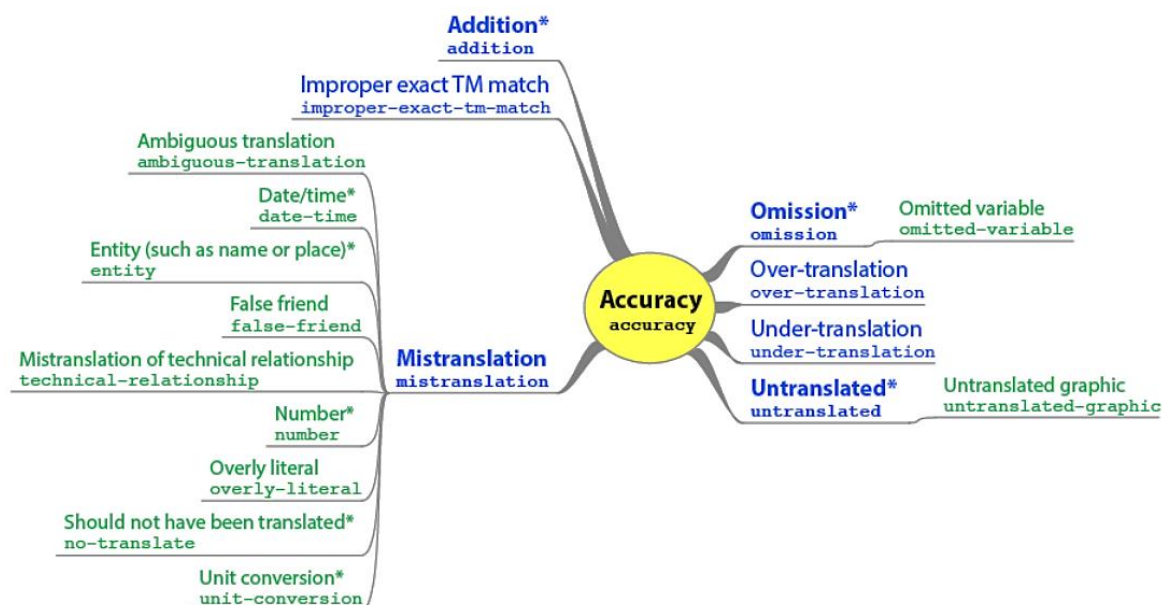


Fig. 3. Hierarchical structure of Multidimensional Quality Metrics dimension of Accuracy (Lommel et al., 2015a)

The presented hierarchy draws attention to the different levels of issue types, i.e., *Accuracy* is the first-level issue, whereas “Mistranslation” is a second-level issue type that contains third-levels issue types such as “Ambiguous translation” or “Mistranslation of technical relationship”. The very important aspect of the metrics to note is that “MQM can be realized at any level of granularity” (Lommel et. al., 2015a), i.e., a researcher can choose the level at which to carry out the content analysis. This metric does not bond to check each and every category present in its hierarchy; it “allows selection of any subset appropriate for the task at hand” (Popović, 2018, p. 136). This logic is based on a fact that in order to create a TQA standard that “adopts the “functionalist” approach that quality can be defined by how well a text meets its communicative purpose” (Lommel et al., 2015b). For this reason, a researcher is free to choose any of the dimensions as well as level issues that fit the most in their analysis, i.e., if only Accuracy and Style were analysed to meet the aim of a research, this MQM metric consisting of two dimensions would be considered valid (Lommel, Uszkoreit & Burchardt, 2014, p. 458; Lommel, 2018, p. 118). Furthermore, it is noted that in frequent cases it is difficult to agree on second-, third- or fourth-level categories and that such ambiguous situations should be solved by using the “parent” levels (Lommel et al., 2014, p. 459; Lommel, 2018, p. 117), e.g., if it is not clear whether an error falls under “Number” or “Unit conversion” category, a higher level category – “Mistranslation” – can be used (see Fig. 3). An attention must also be drawn to the fact that MQM Typology presents a well-defined list of each and every issue with their definitions so that they are used by researchers who decide to employ MQM as TQA tool (more information in Lommel et al. (2015a). This reduces a possible subjectivity and when assigning errors to particular issue types of the metrics.

The relatively simple adaptability of the typology also allows to measure the severity of particular issues. Taking up the ideas present in SAE J2450 and LISA QA models, the MQM model suggests that issues can be of four levels: minor, major, critical and null (Lommel et al., 2015b). Naturally, the

more severe the issue is, the more impact it has on a translation quality. MQM model suggests that severity levels are assigned not to categories, but to individual errors (Lommel, 2018, p. 120); moreover, each severity level has an assigned weight in numerical value that “serves as multipliers for error penalties” (Lommel et al., 2015b). The definitions and weights (indicated in the brackets) of severity levels are as follow:

1. **Null** (0). This level is assigned to issues that are not errors. More specifically, this level is to be assigned to the units that would rather be changed because of preference, or to the units that are systematically repeated and can be easily fixed.
2. **Minor** (1). This level is assigned to issues that do not have negative influence on usability or understandability of the content. This includes the errors that are easily fixed, e.g., extra spaces or full stops, unnecessary capitalisation, slight grammatical errors, etc.
3. **Major** (10). This level is assigned to issues that have negative influence on understandability of the content but does not make it unusable. More specifically, this includes errors that may require more attention of the reader to understand the meaning; therefore, they must be fixed. However, the inclusion of major errors would not result in negative outcomes, such as unfitness of a text for its purpose.
4. **Critical** (100). This level is assigned to issues that make the content unsuitable for use. This includes the errors that changes the meaning of the text and that may cause harm to its user (Lommel et al., 2015b; Lommel, 2018, p. 121).

The severity levels and their weights are used in an MQM metric to evaluate a translation according to the following formula:

$$Score = 1 - \frac{Penalties}{Word\ Count}$$

Here, a translation is deemed completely acceptable in terms of its quality if the score is 1 or 100%. For example, if a translation contains 100 words and 2 minor and 3 major errors (i.e., 32 penalty points), the translation quality score would be 0.68, or 68%. It is also noted that negative scores are possible if a translation contains a lot of errors; therefore, there are more penalty points than words in the text (Lommel, 2018, p. 122). However, it must be noted that very few research studies have been found to employ the mentioned formula to obtain a translation quality score (e.g., Klubička et al. (2018) adjusted the formula to calculate an error ratio; Freitag et al. (2021) adjusted weights of severity levels and calculated a score based on only major, minor and neutral errors).

Furthermore, together with the analytical quality assessment method that has just been discussed, MQM also suggests holistic method to assess the overall characteristics of a translation in terms of reader impression, style, accuracy, usability, etc. (Lommel et al., 2015b). It is recommended to employ the holistic method by questioning evaluators on the high-level issues (i.e., the discussed dimensions) and asking them to rate a translation on a Likert scale or a similar system (Lommel, 2018, p. 123). The holistic method, therefore, allows a researcher to learn whether a translation is deemed as acceptable by its users (Lommel, 2018, p. 123); if it is deemed unacceptable, it paves the way for an analytical method that would be used to analyse the errors that have the greatest impact on a translation’s acceptability.

The analysed TQA systems show that there has been an attempt made towards a creation of one translation quality assessment system that could universally be used to evaluate a product of

translation process. The EU-funded QTLaunchPad project has introduced a very versatile MQM system that allows its users to employ only those dimensions that are needed for their research; it can also be used to analyse either a text or a translation of any language direction. Even though it could be discussed whether this method is suitable for all languages as well as language pairs, the applicability of the method becomes the most appealing characteristic as well as the reason to employ it in evaluation of translation quality.

Taking everything into account, a translation can only be deemed acceptable or unacceptable by its target user which makes it an especially subjective, therefore, complex notion that requires extensive analysis considering the main elements that make a translation acceptable. The analysed research has shown that a translation is deemed acceptable if it meets the needs of its users, and if it contains no significant errors that would disrupt the idea of the text; these needs and errors fall under the three presented elements of acceptability, i.e., usability, quality and satisfaction. Whilst the first and (especially) the last elements of acceptability depend completely on the opinion of a text user, the quality is mainly influenced by the (number of) errors that are / are not present in a translation. For this reason, it is important to analyse the weakest aspects of machine translation output that need to be improved so that machine translation output is deemed acceptable by its users. A number of different errors classification taxonomies have paved a path towards universal TQA systems that allow researchers to base their studies on the same quality assessment methods; a particularly significant Multidimensional Quality Metrics (MQM) typology was introduced in EU funded QTLaunchPad project that adopts the functionalist approach that a translation quality can be analysed only by considering those dimensions of MQM typology that are relevant for a research. Therefore, applying the MQM typology in evaluation of the quality element of acceptability may prove to deliver valid results that would show how acceptability of machine translation output can be improved by considering the errors that cause the most comprehension issues to its user. Having analysed the most important theoretical aspects of acceptability as well as translation quality assessment, the analytical part of the work will be focused on expert and non-expert acceptability of machine translation output and the analysis of its quality.

2. Human acceptability of machine translation

The analytical part of the work analyses experts and non-experts' acceptability of machine translation in terms of satisfaction, usability and quality. Since the results are obtained by presenting surveys to people with and without linguistic background, the extent to which machine translation is deemed acceptable is expected to be contrasting. This part of the work also focuses on the quality element in more detail to obtain insight on the errors that have the most impact on the machine translation quality.

2.1. Methodology

The methods used in the research are descriptive, comparative, and qualitative. The descriptive theoretical analysis is used to review literature on the relevant topic of acceptability and translation quality assessment. The qualitative method is employed to analyse errors present in the translation as well as answers of the online survey, the data of which are also analysed quantitatively. The comparative analysis is applied to compare the results of the online surveys on the acceptability of the machine translation.

It must first be noted that the source text of the analysed translation was taken from "BBC" online news website. Since the translation was to be handed out to translators and general public, a simple and random yet relevant article concerning the topic of COVID-19 was chosen (see Appendix 1). The freely available Google Neural Machine Translation System was chosen because of its popularity and ease of access; it was employed to translate the English source text into Lithuanian. The produced translation contains 532 words (see both source and target texts in Appendix 1). It must also be noted that the research analyses human acceptability of machine translation in general and not acceptability of machine translation produced via the Google Neural Machine Translation System.

The analysis of the translation was twofold. First, a holistic analysis of acceptability by machine translation users was performed. The total number of participants of the study was 90, all of which are native speakers of Lithuanian. 20 of them are students of first- and second-year translation and localisation master studies who have experience in translation and/or a degree in translation or languages. These participants were named *Experts* (i.e., "senior translators or trained linguists (Castilho et al., 2018, p. 23). 70 other participants are members of general public to whom a survey was presented in the social media network "Facebook". The survey remained open for two weeks. The participants were asked to dedicate 5 minutes of their time to read a short translation and fill in the survey on the acceptability of the machine translation. The survey was anonymous to receive as many results as possible. Thus, the age, gender as well as the background of the participants are unknown. These 70 participants were named *Non-Experts*. Different experiment participants were chosen to find out whether machine translation output in Lithuanian as a target language is accepted in the same way by basic users of machine translation and translators.

The acceptability of the translation was evaluated by experts and non-experts according to the acceptability model proposed by Castilho (2016) in which the three main components are usability, quality and satisfaction (p. 58). Participants were introduced with two online surveys presented in their native – Lithuanian – language: one for experts (see Appendix 2) and one for non-experts (see Appendix 3). In the first part of both surveys, the participants were asked to read a short translation excerpt consisting of seven sentences (114 words). The participants were not presented with a full translation words to obtain as many results as possible as the presentation of the full translation would risk the respondents' interest in the task.

Both experts and non-experts were asked to read a short excerpt consisting of seven sentences and evaluate six short statements on the Likert scale (see Appendix 2 and Appendix 3), to which values from 1 to 5 were assigned (1 = Disagree, 2 = Somewhat disagree, 3 = Neither agree nor disagree, 4 = Somewhat agree, 5 = Agree). Statements 1 and 2 were added to measure satisfaction; statements 3 and 4 were added to measure usability; statements 5 and 6 were added to measure quality. The results of this part were then systemised using MS Excel 2016 by gathering them into a table and preparing them for a comparative one-way MANOVA analysis with IBM SPSS Statistics 22.0 to obtain general insight on the machine translation's acceptability. The results were then analysed in more detail by discussing the answers of each and every statement separately.

The second part of the surveys was different for experts and non-experts. Experts were asked to read every sentence and indicate if there was an error in it, to copy and paste the error if there was any, to answer whether they would correct the mistake, and to identify the severity of the error, i.e., critical, major, minor, null (see Appendix 2). The meaning of each error severity was provided to experts prior to the experiment. In the meantime, non-experts were asked to read every sentence and indicate whether it was comprehensible, partially comprehensible or incomprehensible. If the sentence was partially comprehensible or incomprehensible, non-experts were asked to indicate the segment that causes comprehension issues (see Appendix 3). The presentation of different second part of the surveys to experts and non-experts was chosen to avoid possible deviations due to uneven understanding of the severity levels that could not be sufficiently communicated to each non-expert participating in the survey.

Moreover, the reason behind different parts of the survey for experts and non-experts is directly linked with the second part of the study in which the overall quality of the machine translation was thoroughly analysed. The machine translation quality was assessed by employing the MQM typology, more specifically – its dimensions of Accuracy, Fluency, Style and Locale convention. The dimensions of Compatibility, Design, Terminology and Verity were not taken into account as the source and its target text (see Appendix 1) are designated to simply provide general knowledge to public; the dimension of Internationalisation was not analysed as well as this branch applies only to the source text. Therefore, analysing the details of the texts' design, internationalisation possibilities, specific terminology usage problems as well as verity were not relevant.

The analysed types of issues, if there are any, of specific dimensions are as follow:

- Accuracy dimension. Issue types:
 - Addition
 - Improper exact TM match
 - Mistranslation
 - Omission
 - Over-translation
 - Under-translation
 - Untranslated
- Fluency dimension. Issue types:
 - Grammar
 - Grammatical register
 - Inconsistency
 - Spelling
 - Typography
 - Unintelligible
 - Ambiguity
 - Character encoding
 - Coherence
 - Cohesion
 - Corpus conformance
 - Duplication
 - Index / TOC
 - Link / cross-reference

- Nonallowed characters
- Offensive
- Style dimension. Issue types:
 - Register
 - Awkward
 - Company style
- Locale convention dimension. Issue types:
 - Address format
 - Calendar type
 - Currency format
 - Date format
 - Locale-specific punctuation
- Pattern problem
- Sorting
- Inconsistent style
- Third-party style
- Unidiomatic
- Measurement format
- Name format
- Shortcut key
- Telephone format
- Time format

The examples are presented in full sentences in an *italic* font by highlighting the issues of accuracy, fluency, style or locale convention found in the target sentence in a ***bold italic*** font. If only a segment or word is presented, it is marked in *italic* font. The full list of the dimensions as well as every issue types' definitions will not be presented for the sake of scope of this paper. However, the extensive list can be found in Lommel et al. (2015b). These definitions are used to describe the errors that are found in the machine produced translation (see Appendix 1).

Moreover, the severity levels of errors as well as their weights are assigned by employing the MQM model (Lommel et al., 2015b). These are:

- Critical (100)
- Major (10)
- Minor (1)
- Null (0).

The weights are then used to evaluate the overall translation quality score according to the following formula:

$$Score = 1 - \frac{Penalties}{Word\ Count}$$

The overall quality score as well as the severity levels are then compared to the online surveys' results. The indicated issues and their severity levels as assigned by experts as well as their opinion on whether the issue should be fixed is compared with the non-experts' answers on the comprehensibility of the sentences as well as the indicated issues, if there are any. The comparative analysis of the results obtained by analysing the errors as well as their impact on the overall quality of the machine produced translation allow to get insight on which types of errors have the biggest impact on translation quality – therefore, acceptability of machine translation output – as perceived by experts, i.e., individuals who are oriented towards the improvement of machine translation systems, and as perceived by non-experts, i.e., daily users of machine translation systems. This set of different analysis methods allows not only to draw conclusions on overall quality of machine translation as well as its acceptability, but also to find out which error types seem to be the most unacceptable for experts and whether these error types have same importance to basic users.

The analysis of machine translation acceptability as perceived by experts and non-experts and the machine translation quality assessment allowed to draw conclusions not only on the possibly different acceptability of the machine translation as evaluated by basic users and linguists, but also allows to

get insight into the quality aspects that have significant impact on the comprehensibility of the translation and text in overall. The multifaceted study, therefore, provided valuable results based on which neural machine translation systems may be improved and may also act as exemplary guidelines for content creators as it identifies the shortcomings of a text that cause the most comprehension issues to its target user.

2.2. Acceptability of machine translation as assessed by experts and non-experts

The results of the surveys on acceptability presented to experts and non-experts in their native – Lithuanian – language first and foremost draw attention to the different values of Likert scale assigned by participants for the six presented statements intended for the evaluation of acceptability elements. The statements are as follow:

1. The main idea of the text is clear (Lit. Pagrindinė teksto mintis yra aiški).
2. The text is written in an orderly, understandable way (Lit. Tekstas tvarkingas, suprantamas).
3. The text is useful, informative (Lit. Šis tekstas yra naudingas, informatyvus).
4. I would trust the source of the text (Lit. Pasitikėčiau šaltiniu, kuriame šis tekstas yra pateiktas).
5. The text sounds natural (Lit. Tekstas skamba natūraliai).
6. The text is correct, and its quality is excellent (Lit. Tekstas taisyklingas, jo kokybė puiki).

The results of the surveys show that the general opinion on the acceptability does not significantly differ among experts and non-experts; these results are presented in **Table 1**.

Table 1. One-way MANOVA results on machine translation acceptability

–	Participant	Mean	Std. Deviation
Statement 1	Expert	2.80	1.005
	Non-expert	2.66	1.239
	Total	2.70	1.172
Statement 2	Expert	1.50	.513
	Non-expert	1.80	.926
	Total	1.71	.837
Statement 3	Expert	2.45	.826
	Non-expert	2.42	1.108
	Total	2.43	1.030
Statement 4	Expert	1.45	.686
	Non-expert	1.60	.670
	Total	1.56	.673
Statement 5	Expert	1.20	.410
	Non-expert	1.50	.678
	Total	1.41	.625
Statement 6	Expert	1.20	.410
	Non-expert	1.60	.728
	Total	1.49	.676

The results obtained by applying a one-way MANOVA (see Table 1) suggest that the machine translation is deemed rather unacceptable by both experts and non-experts as neither of the statements reached an average of assigned value 3 (i.e., the assigned values show that the majority of the participants marked “Somewhat disagree” (2) or “Disagree” (1) for the majority of the statements); however, the translation cannot be confidently deemed unacceptable considering the standard deviations that are apparent in the results of every statement. The standard deviation was the least significant for the statements 5 and 6 that were presented to assess quality of the machine translation which showed that experts and non-experts together and separately disagree that the quality of the text is high; therefore, considering similar deviations as well as coinciding mean result of the values assigned, conclusions could be drawn that the presented machine translation’s quality is low. The results of statement 1 and 4 that accordingly were meant to assess the satisfaction and usability of the text show that experts and non-experts shared the same views once again as the results of the two statements have very minor differences in terms of the mean value; however, the high score of standard deviation of the statement 1 shows that both experts and non-experts share uneven views on the satisfaction with the machine translation. Moreover, the results of statements 2 (presented to assess satisfaction) and 3 (presented to assess usability) provide conflicting results as well. Although the expert and non-experts’ means do not show any significant differences, standard deviations of the two groups’ results indicate that experts show more consensus in the evaluation of statements 2 and 3 while non-experts share more diverse views. For this reason, strong conclusions on the acceptability of the machine translation cannot be made as experts and non-experts do not agree on the satisfaction with and usability of it despite the unanimous opinion on the low quality of the text; for this reason, it is important to analyse the results of the survey on satisfaction, usability and quality separately and in greater detail.

2.2.1. Expert and non-expert satisfaction with the machine translation

The statements 1 and 2 of the first part of the surveys were provided to experts and non-experts to analyse the overall satisfaction with the presented machine translation. The results are presented in the Fig. 4 and Fig. 5 below.

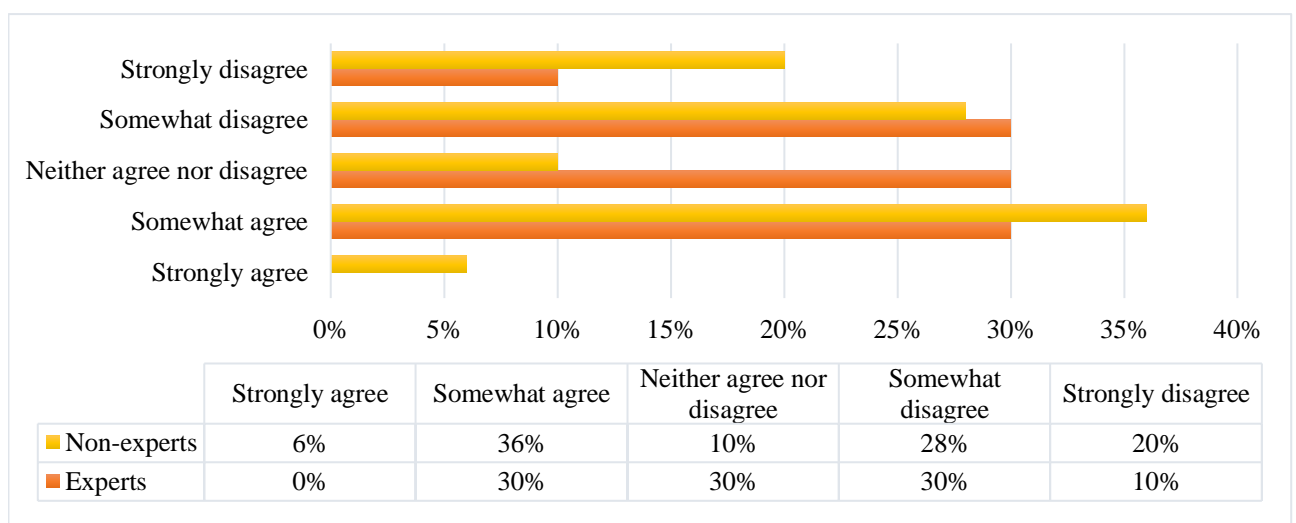


Fig. 4. Answers to statement 1: The main idea of the text is clear

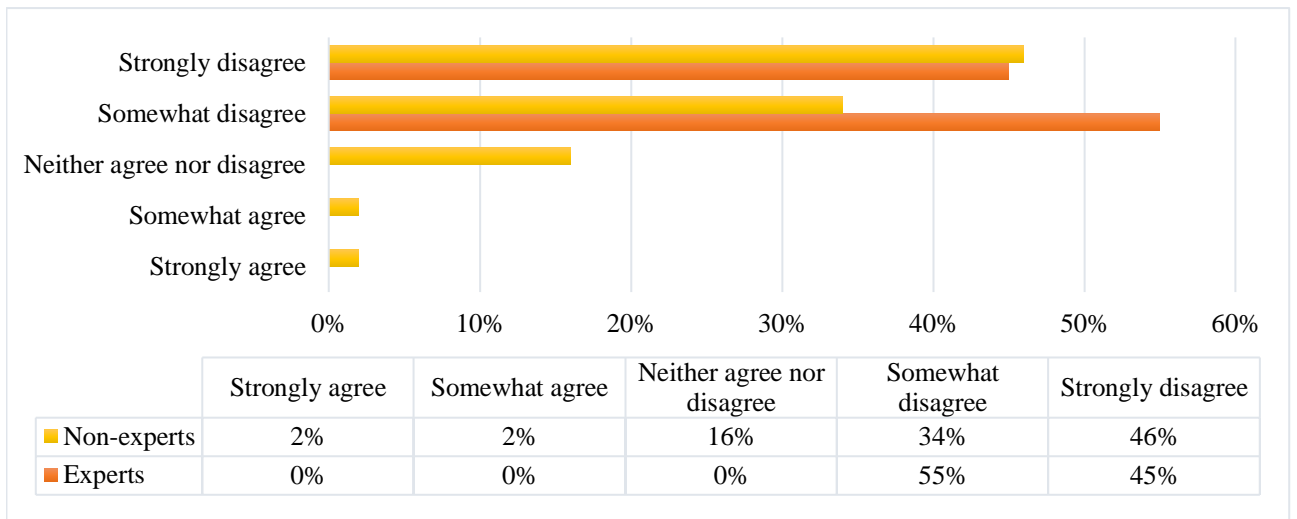


Fig. 5. Answers to statement 2: The text is written in an orderly, understandable way

The results show that experts' and non-experts' opinion in terms of satisfaction differs as whilst non-experts (36% of them) are mainly satisfied with the way the main idea of the source text is rendered with the use of the freely available machine translation system, experts remain more sceptical and disagree on a single value (see Fig. 4). However, the results of the statement 2 (see Fig. 5) show that experts and non-experts share the same views on manner that the text is presented in as both of the groups (100% of experts and 80% of non-experts) are prone to somewhat disagree or disagree that the text is written in an orderly and understandable way. Therefore, the results of the first two statements that are intended to evaluate the first element of acceptability – satisfaction – show that even though experts and non-experts agree that the manner in which the translation is rendered is poor, non-experts show satisfaction towards the presentation of the main idea of the text, whereas experts disagree on this question. This disagreement is also apparent in the high number of the standard deviation (see Table 1). Taking everything into account, even though neither experts nor non-experts indicated that the machine translation is satisfactory, non-experts showed more satisfaction with the text than experts who agree that neither the idea of the text is clear, nor the text is written in a satisfactory manner.

2.2.2. Expert and non-expert opinion on the machine translation usability

The third and the fourth statements of the first part of the surveys (see Appendices 2 & 3) were provided to analyse the usability of the machine translation. The results are presented in Fig. 6 and Fig. 7.

Non-experts and experts' opinion on the usability of the translation differ as well. The results of the third statement (see Fig. 6) show 50% of experts neither agree nor disagree that the translation is useful and informative, whereas the opinion of non-experts varies (e.g., 24% somewhat agree that the translation is useful and informative, whilst 34% somewhat disagree and 24% disagree). The variation of the results among non-experts might arise from the fact that the participants were not provided with instructions as well as definitions of what is deemed useful and what is not; therefore, the respondents might have had different understanding of the notion. Also, the results of the third statement that measures usability opposes the good opinion on satisfaction of non-experts who agree that the idea of the text is clear (see Fig. 4); however, they correspond to the results of satisfaction on opinion on the manner in which the text is rendered as well as its comprehensibility (see Fig. 5). This

signalises that non-experts deem a translation useful and informative not only if the idea of it is clear; the way in which this idea is presented also has a great impact on the non-experts understanding of a translation’s usability.

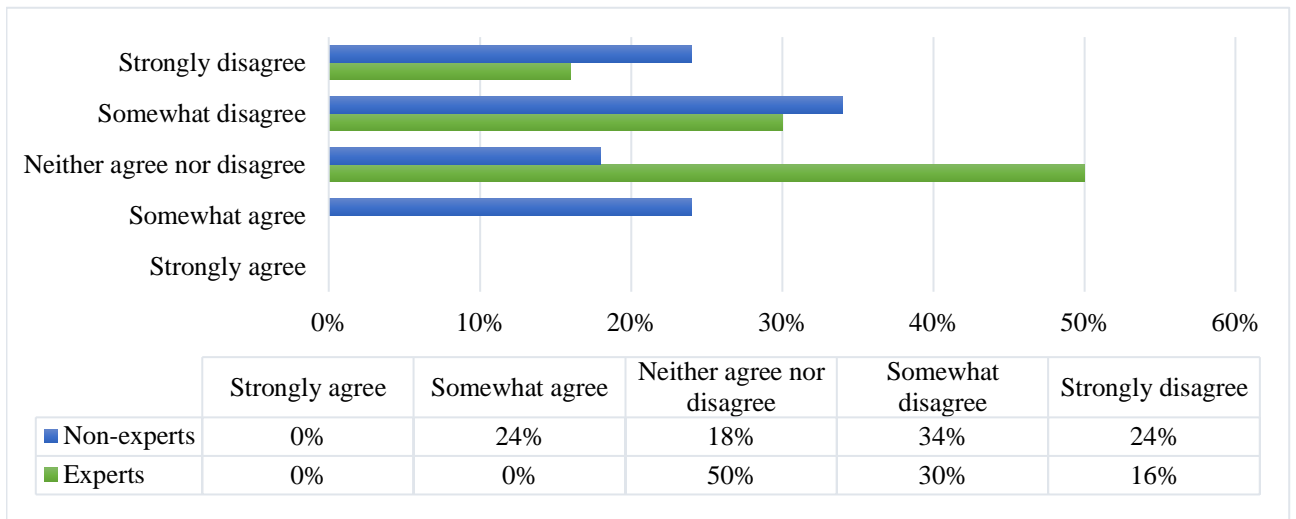


Fig. 6. Answers to statement 3: The text is useful, informative

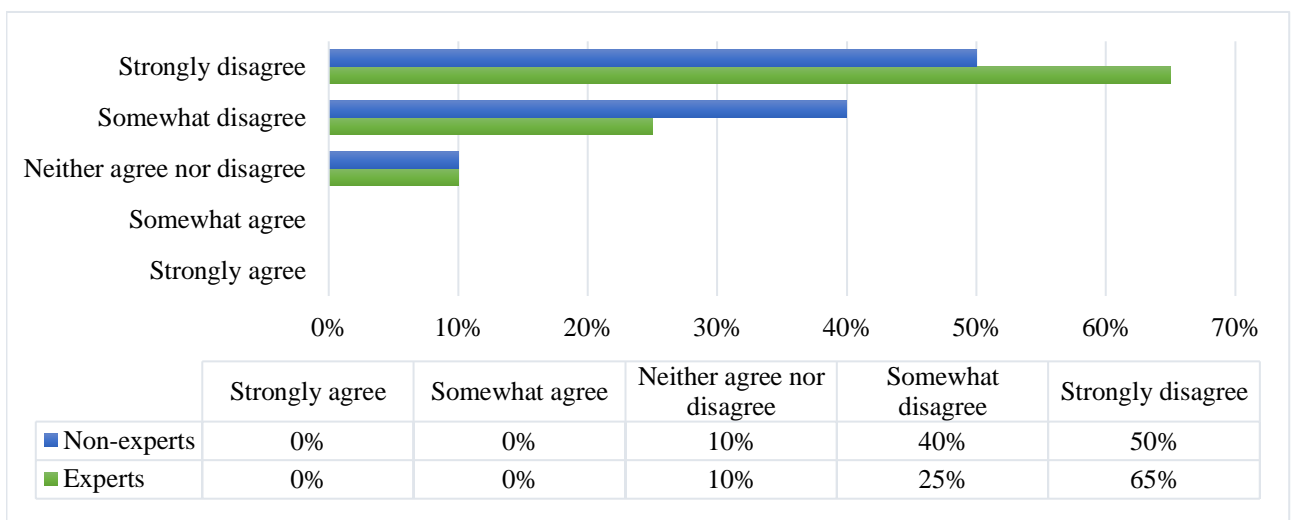


Fig. 7. Answers to statement 4: I would trust the source of the text

The results of the fourth statement (see Fig. 7) show whether the participants would be apt to use the website or any other source in which such texts would be presented. Both experts and non-experts’ opinion on this aspect coincides – neither of the participants agree that they would trust the source of the text as the majority of them (50% of non-experts and 65% of experts) disagrees with the provided statement and does not view the text and its source to be reliable; it thus reduces the overall score of usability of the machine translation.

In general, the results on the usability of the machine translation show that the opinion of the experts and non-experts differ. Whilst experts neither agree nor disagree that the translation is useful and informative, they strongly admit that the source of such a text is not usable. On the other hand, even though non-experts’ views coincide with the experts’ on the reliability of the source, non-experts seem to be more critical towards the usefulness and informativity of the text despite the fact that they

are satisfied and agree that the main idea of the text is clear (as seen in Fig. 4). Therefore, it could be said that non-experts' views once again are conflicting which could have resulted because of the absence of the usability definition, whilst experts share unanimous opinion that the machine translation cannot be deemed usable.

2.2.3. Expert and non-expert opinion on the machine translation quality

The fifth and the sixth statements of the first part of the surveys (see Appendices 2 & 3) were provided to analyse the quality of the machine translation as perceived by the experiment participants. The results are presented in the figures below.

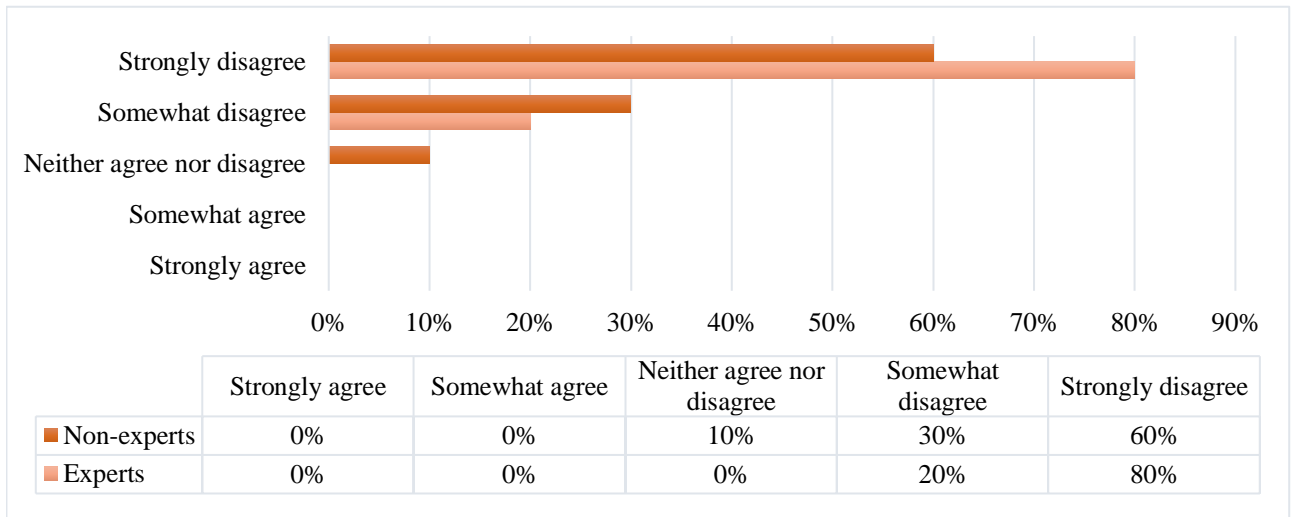


Fig. 8. Answers to statement 5: The text sounds natural

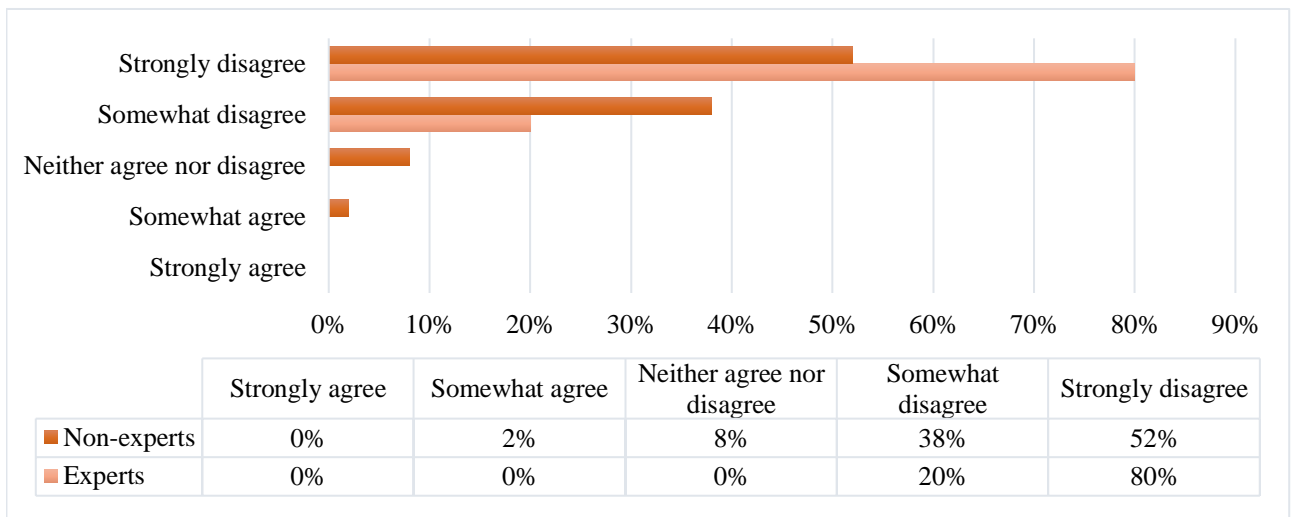


Fig. 9. Answers to statement 6: The text is correct, and its quality is excellent

Although experts and non-experts shared different views on the machine translation usability and satisfaction, they seemed to agree on its quality. The majority of experts (80%) and non-experts (60%) disagree that the text sounds natural (see Fig. 8); they also have the same opinion on the text correctness and its quality in overall (80% of experts and 52% of non-experts disagree, and 20% of experts and 38% of non-experts somewhat disagree with the statement 6; see Fig. 9). The results of

the statements 5 and 6, the mean values as well as deviations of those (see Table 1) show that the vast majority of the respondents assert that the quality of the presented text is low. This makes the 3rd element – quality – the least acceptable factor that has the most influence on the overall acceptability of the machine translation as views of experts and non-experts coincide.

Taking everything into account, the machine translation of a general text taken from a news website is deemed unacceptable by both experts and non-experts. The results show that neither of the statements have resulted in an average of 3 or higher which would indicate that the translation is neither good nor bad; this indicates that all three criteria of acceptability – satisfaction, usability, quality – are not met. Nevertheless, the results showed that non-experts have a more optimistic opinion on the text’s usability and satisfaction as they assigned higher values to certain aspects of the text than experts. However, the opinion of experts and non-experts coincided when evaluating the quality of the text. The majority of the respondents assigned low values to evaluate certain aspects of the text which showed that the quality of machine translation is the main reason of the unacceptability of it as both experts and non-experts agreed that the issues are apparent. This draws attention to the fact that machine translation is still flawed and that many improvements should be made to make it acceptable to its target users. Nonetheless, it can be assumed that certain errors caused more comprehension issues to both experts and non-experts than others; therefore, they have more impact on the overall acceptability of machine translation. This is to be found out in the next sections of the thesis.

2.3. Analysis of machine translation errors

The surveys’ results on machine translation acceptability showed that the most problematic aspect of machine translation that both experts and non-experts agree on was its quality. For this reason, a thorough analysis of the machine translation was carried out (see Appendix 4) to find out what issues have the most influence on it.

The analysis of the machine translation of a general source text retrieved from a widely used news website by employing the MQM error classification and analysing the Accuracy, Fluency, Style and Locale convention dimensions showed that the overall number of the found errors in the translation was 80 (see Fig. 10).

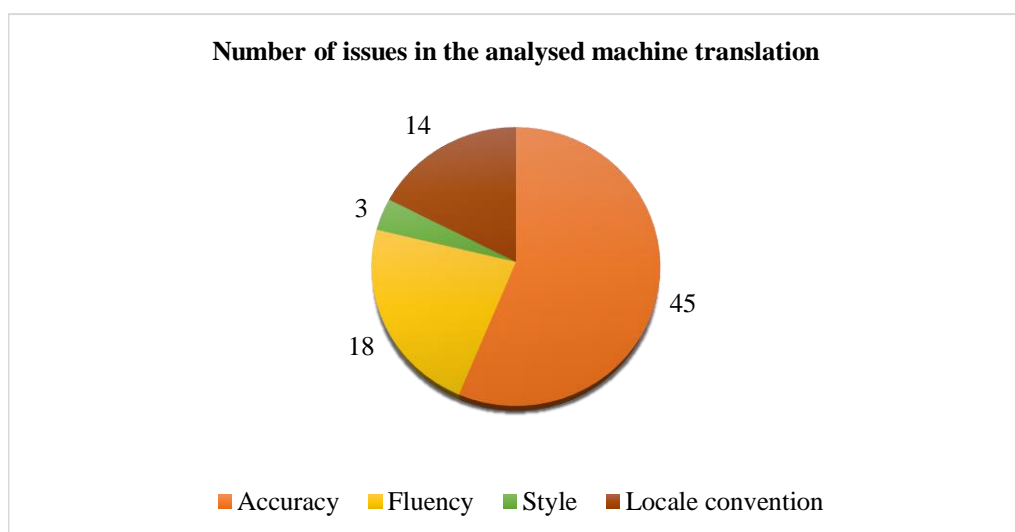


Fig. 10. Number of issues in the analysed machine translation

Moreover, the analysis of the errors and their weights (see Appendix 4) showed that the overall translation quality score of the translation is negative: $1 - \frac{726}{532} = -0.364$, or -36% .

The considerably big number of issues found in a short translation consisting of 532 words as well as the significantly low-quality score of -36% signals that the majority of the translated segments contain serious flaws that allows to draw a conclusion that the quality of machine translation is far from perfect. The results showed that the most frequent errors are related to the accuracy of the translation (45 found issues) followed by fluency (18 found issues), locale convention (14 found issues) and style (3 found issues) errors (see Fig. 10). The errors of distinct dimensions and their weights are analysed further below.

Accuracy dimension

The results of the produced machine translation analysis draw an exceptional attention to the issues of the Accuracy dimension as the number of those is the highest. Making up of 56% of all of the found mistakes in the machine translation, accuracy issues seem to cause the most comprehension issues. The found second-level errors of the Accuracy dimension are shown in Fig. 11.

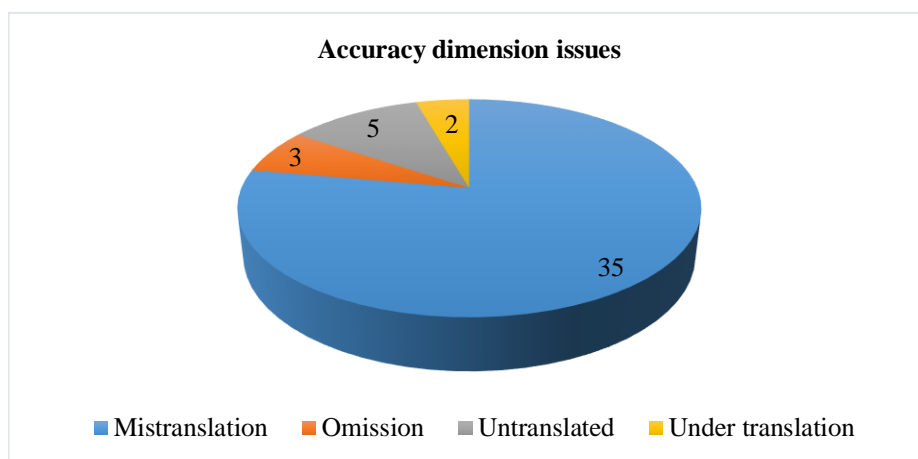


Fig. 11. Issues of the Accuracy dimension

The most common second-level errors of the Accuracy dimension are mistranslation with the count of 35 out of 45. For example:

- Source (EN): Covid vaccine: Germany urged to back AstraZeneca **jab** for over-65s
Target (LT): „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ **žandikaulį** vyresniems nei 65 metų žmonėms
- Source (EN): Carsten Watzl urged Angela Merkel **to have the vaccine live on TV** to prove it is safe.
Target (LT): Carstenas Watzlas paragino Angelą Merkel **vakciną tiesiogiai transliuoti per televiziją**, kad įrodytų, jog ji yra saugi.
- Source (EN): **But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.**
Target (LT): Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, **rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.**

The other relatively few errors are caused by omitted, under translated or untranslated segments. The examples are presented accordingly:

4. Source (EN): *In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.*
Target (LT): *BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių.*
5. Source (EN): *AstraZeneca itself says the vaccine is effective at all age groups.*
Target (LT): *Pati „AstraZeneca“ teigia, kad vakcina yra veiksminga bet kokio amžiaus.*
6. Source (EN): *The UK is among countries that have approved the jab for all age groups.*
Target (LT): *JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jab.*

In overall, the penalties score of the Accuracy dimension is 699, which corresponds to 96% of the number of penalties of the errors found in the machine translation (see Appendix 4). This draws attention to the fact that accuracy is the main issue and reason of a low-quality machine translation output that is mainly caused by mistranslations that make the translation unusable. Even though the majority of the sentences include accuracy null, minor or major errors (e.g., see example 2 for the mistranslation in the segment *vakciną tiesiogiai transliuoti per televiziją* (Eng. to have the vaccine live on TV), see example 5 for the under translation in the segment in the segment *bet kokio amžiaus* (Eng. at all age groups) that do not make the translation unusable, the six found critical errors, one of which might even have negative influence on human life (see example 3; the absence of the negation “not” in the mistranslated segment *rekomendavo jo naudoti vyresniems nei 65 metų* completely changes the idea of the sentence) show that critical accuracy errors have the biggest weight on the translation quality. It, therefore, can be stated that critical accuracy errors are the main cause for the low machine translation quality as they significantly alter the idea of the text.

Fluency dimension

Other issues that have negative influence over the translation quality is of the Fluency dimension. Making up of 22.5% of all of the found mistakes in the machine translation, fluency issues come send in terms of quantity of errors found in the machine translation. The found second-level issues of the Fluency dimension are shown in Fig. 12.

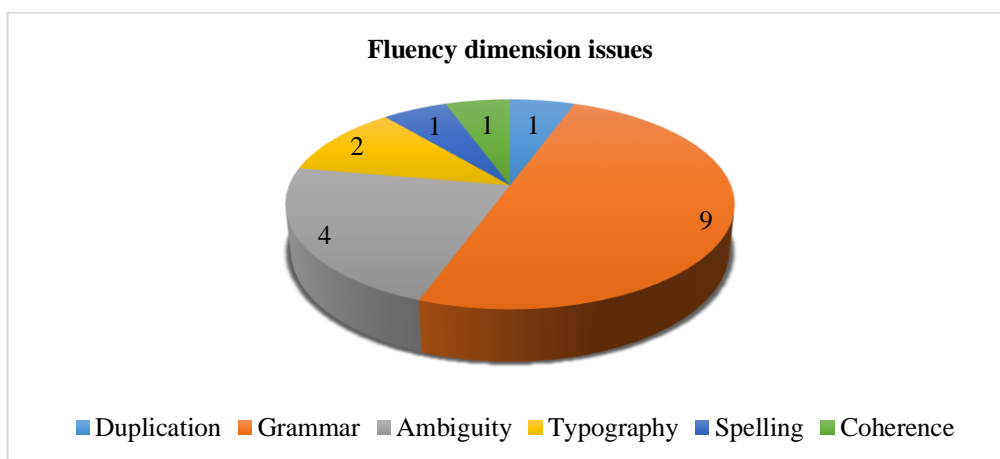


Fig. 12. Issues of the Fluency dimension

The most common second-level subtype of the Fluency dimension is grammar with the count of 9 out of 18. For example:

7. Source (EN): *The recent data from Scotland clearly show it elicits an immune response, the elderly are protected from severe disease by this vaccine.*
Target (LT): *Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.* – word order.
8. Source (EN): *Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.*
Target (LT): *Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų banga, jei per greitai bus panaikinta užraktas.* – word form.

The other relatively few errors are caused by duplication, ambiguity, typography, spelling or coherence issues, for instance:

9. Source (EN): *The EU's medical regulator approved the use of the Oxford-AstraZeneca vaccine for all age groups in January.*
Target (LT): *ES medicinos priežiūros institucija sausio mėn. **Patvirtino** „Oxford-AstraZeneca“ vakciną naudojimą visoms amžiaus grupėms.* – spelling (capitalisation).
10. Source (EN): *But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.*
Target (LT): *Jie nurodė nepakankamus duomenis apie **jo** veiksmingumą vyresnio amžiaus žmonėms.* – ambiguity (unclear reference).
11. Source (EN): *On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated recommendation" on the AstraZeneca vaccine.*
Target (LT): *Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą, AstraZeneca“ vakciną rekomendaciją.* – typography.

The penalties score of the Fluency dimension is 21, which corresponds to 2.9% of the overall number of penalties of the errors found in the machine translation (see Appendix 4). Whilst the score is significantly lower than that of the Accuracy dimension, the found errors of grammar, spelling, typography and other still may cause comprehension issues due to unnaturalness or ambiguity of the machine translation output. The presence of foreign elements apparent in small details of the segments (for example, see example 11) have a significant impact on the acceptability of the machine translation as the mentioned and found unnatural word orders, word forms, spellings, quotation marks and other reveal that the presented text is not designed for a specific locale only; therefore, even though there is no critical issues of the Fluency dimension found after analysing the machine translation, the presence of foreign elements as well as the lack of coherency make a translation unacceptable for the Lithuanian locale.

Style dimension

The analysis of the errors found in the machine translation also shows some issues in the style. Making up only 3.8% of all of the found mistakes in the machine translation, there seems to be no significant problems in the manner in which the translation is rendered. All 3 of the found errors of the Style dimension are of the “unidiomatic” 2nd level issue type and are as follow:

12. Source (EN): *"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.*

Target (LT): *"- Manau, kad Vokietija taip pat greitai pakeis kursą"- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.*

13. Source (EN): *Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.*

Target (LT): *Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atveju banga, jei per greitai bus panaikinta užraktas.*

The null (0) weights are assigned to the style issues (see Appendix 4) as the manner in which the translations of the segments is rendered could be edited by preference; however, it should be noted that the target text contains unidiomatic segments as compared with the source text. The style of the sentences and the lack of idiomatic expressions in them are also directly linked with the accuracy issues, the majority of which are noted as 3rd level issue – overly literal translation. This too literal translation also has a great impact on the style of the translation; thus, the idiomaticity of the source text is affected as well. However, the unidiomatic style of the sentences does not make them unusable. For this reason, style issues do not have significant influence over the machine translation quality.

Locale convention dimension

A different set of issues found in the machine translation fall under the Locale convention dimension that concerns the compliance with specific locale expectations. Making up of 17.5% of all of the found mistakes in the machine translation of the news article, locale convention issues are the third most common issues after the already-analysed accuracy and fluency issues. The found second-level errors of the Locale convention dimension are presented in Fig. 13.

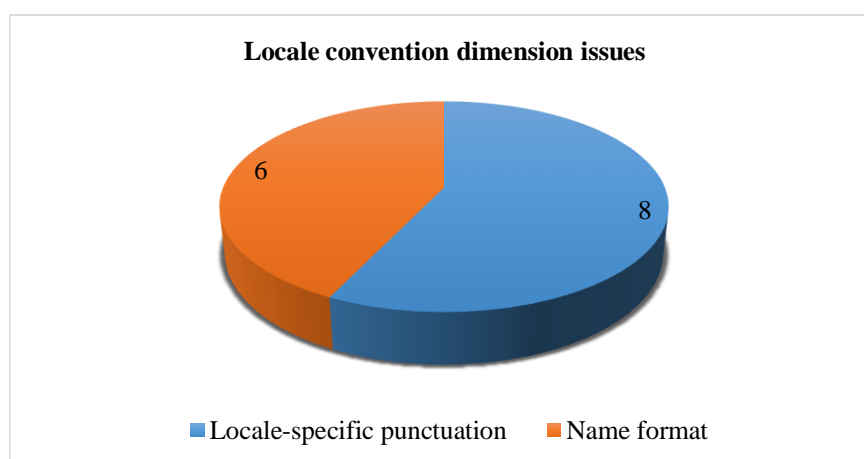


Fig. 13. Issues of the Locale convention dimension

The most common second-level subtype of the Locale convention dimension is locale-specific punctuation with the count of 8 out of 14. The examples are:

14. Source (EN): *"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.*

Target (LT): *"- Manau, kad Vokietija taip pat greitai pakeis kursą"- sakė prof. Watzlas **BBC** radijo 4 laidos "Broadcasting House" programai.*

15. Source (EN): *"If at that point [Chancellor] Angela Merkel were to go on live television and have the vaccine, that would be great," he added.*

Target (LT): *„Jei tuo metu [kanclerė] Angela Merkel eitų į tiesioginę televiziją ir turėtų vakciną, būtų puiku“, - pridūrė jis.*

The rest of the locale convention issues are incorrect name formats. For instance:

16. Source (EN): *Carsten Watzl urged Angela Merkel to have the vaccine live on TV to prove it is safe.*

Target (LT): ***Carstenas Watzlas** paragino Angėlą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.*

17. Source (EN): *In January, French President Emmanuel Macron said the jab was "quasi-ineffective" for older age groups - a claim strongly refuted by the UK government and British medical regulators.*

Target (LT): *Sausio mėn. Prancūzijos prezidentas **Emmanuelis Macronas** teigė, kad vyresnio amžiaus žmonėms šiukšlės buvo „beveik neveiksmingos“ - JK vyriausybė ir Didžiosios Britanijos medicinos priežiūros institucijos tvirtai paneigė šį teiginį.*

The penalties' score of the Locale convention dimension is 6, which corresponds to 0.8% of the overall number of penalties of the errors found in the machine translation (see Appendix 4). Even though the score is considerably low as the locale convention issues do not have a significant impact on the machine translation quality, the 14 found issues show that the chosen machine translation system is not yet adapted for the Lithuanian locale. The usage of incorrect quotation marks as seen in the example 14, the usage of the hyphen instead of the dash after the direct quotation as seen in the example 15, and the usage of letters not present in the Lithuanian alphabet or a lack of phonetic adaptation (examples 16 and 17) do not make the machine translation unusable or unacceptable; however, these small elements have an impact on the overall translation score and they should not be overlooked to achieve the best quality possible.

The analysis of the errors found in the machine translation of the news article retrieved from BBC first and foremost draws attention to the number of translation issues. The short machine translation consisting of 532 words has 80 errors in it that cause comprehension issues in one way or another. The significantly low translation quality score of – 0.362 or –36% that is based on the number of words in the translation and the weights of the 80 errors shows that there are critical errors that have notable impact on the overall comprehensibility of the text. It has been found that the majority of the 80 errors fall under the accuracy branch; the number of assigned penalties is the highest as well because of significant mistranslations of the source that change the meaning of the text, such as the absence of the adverb “not” in the translation that completely changes the direction of the text and may possibly cause harm to human life if it is to be used. The errors of accuracy are followed by issues in fluency, the majority of which are related with such grammatical errors as word form or word order; even though such errors do not have a significant influence over the translation quality score, they disturb the overall flow of the text. Moreover, the translation lacks the idiomatic style that is found in the source text (but in no way the translation's quality is reduced as the idea of the source text remains unchanged); however, it must be noted that this lack of idiomaticity is directly linked with the mistranslation issues that fall under the accuracy branch. More specifically, the simple style of the target text is caused by the overly literal machine translation that has the biggest influence on the low translation quality score. Furthermore, the analysed machine translation also contains issues

that arise due to elements not tailored towards a specific – in this case, Lithuanian – locale that are apparent in the usage of incorrect quotation marks, hyphens and dashes, and incorrect translation of English or French names that include letters absent in the Lithuanian alphabet, or the letters that are not adapted phonetically. Taking everything into account, it could be said that the overall machine translation quality is low and that the main factor of this negative indicator is issues related to the accurate rendering of the information present in the source text. The low translation quality score also coincides with the results of experts and non-experts' acceptability of the machine translation in terms of quality as seen in the previous section of the work; however, the error analysis does not allow to draw conclusions on the types of errors that have the most impact on the acceptability of the machine translation as it acts only as a complementary analysis method in the evaluation of human acceptability of machine translation. For this reason, it is important to conduct an analysis on the relationship between human acceptability of machine translation and errors present in it, and to note the interdependence of the two.

2.4. Influence of produced errors on machine translation acceptability

The previous sections of the analytical part of the work showed that human acceptability of machine translation is low in terms of satisfaction, usability and quality. Even though experts and non-experts show some disagreement on the satisfaction with and the usability of the machine translation as non-experts view the elements of satisfaction and usability as more acceptable than experts, all of the experiment participants agree that the quality of it is significantly poor. The carried-out machine translation errors analysis shows that the translation quality is considerably low; the negative translation score of – 0.362 mainly depends on the critical accuracy errors caused by mistranslation. However, a firm conclusion cannot be made that human acceptability of machine translation is low because of the low translation quality score as the theoretical part of the work has shown that some errors may have a bigger influence over the acceptability of a machine translation in terms of quality than others. For this reason, it is important to find out which factors, i.e., errors make a machine translation unacceptable as perceived by basic users of the machine translation system as well as professionals.

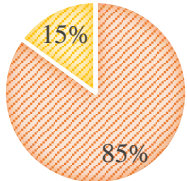
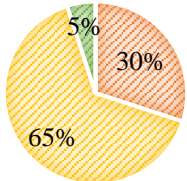
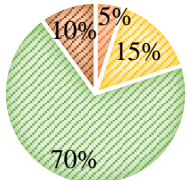
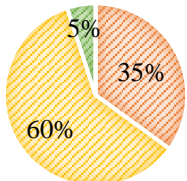
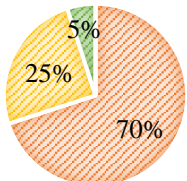
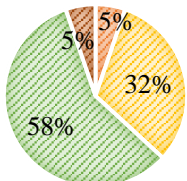
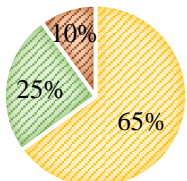
2.4.1. Relationship between errors and experts' acceptability of machine translation

To find out the error types that cause the most comprehension issues to experts, a short excerpt consisting of seven sentences from the analysed machine translation of the news article on COVID-19 topic taken from the news website BBC was analysed. The short excerpt was chosen to keep the respondents' interest in the task as the loss of it would have caused significant deviations or would make the respondents close the survey before finishing it (see p. 17 on the discussion of surveys preparation). The short excerpt contained issues of all of the four analysed dimensions; therefore, there were issues of accuracy, fluency, style and locale convention.

The results of the survey presented to experts first and foremost allows to obtain a general view on how experts assess the presented machine translation quality. The analysis of the acceptability showed that experts agree on the low quality of the text; this is also apparent in the results of the survey (see Table 1, Fig. 8 & Fig. 9).

The majority of the experts (92.5%) identified translation errors in each and every presented sentence. The results are presented in Table 2.

Table 2. Experts' evaluation of machine translation

–	Translated sentence	Error indicated by an expert	Severity of the error as assigned by an expert
1	„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.	žandikaulį	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
2	Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.	vakciną tiesiogiai transliuoti	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
3	ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinos naudojimą visoms amžiaus grupėms.	mėn. Patvirtino	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
4	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	1. rekomendavo jo naudoti 2. į skleidimą	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
5	JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jаб.	jаб	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
6	„Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.“	apsaugo nuo sunkių ligų ši vakcina	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null
7	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.	1. Profas 2. reguliuotojai	 <ul style="list-style-type: none"> ■ Critical ■ Major ■ Minor ■ Null

- 1st sentence: 100% of experts identified the **accuracy error** by noting the word *žandikaulį* (Eng. jab) and assigning the critical (85% of experts) and major (15% of experts) severity level to the error. However, neither of them noted the accuracy error in the word *paremti* (Eng. to back) which was overly translated from the source text.
- 2nd sentence: 100% of experts identified the **accuracy error** by noting the segment *vakciną tiesiogiai transliuoti* (Eng. to have the vaccine live on) and assigned the critical (30%), major (65%) and minor (5%) severity levels to the error. However, the attention must be drawn to the fact that none of the experts have noted the error of locale convention present in the sentence. The name *Carstenas Watzlas* (Eng. Carsten Watzl) was not adapted to meet the standards of the Lithuanian language.
- 3rd sentence: 85% of experts identified the **fluency error** by noting the segment *mėn. Patvirtino* (Eng. approved the use of ... in January) and assigned the critical (5%), major (15%), minor (70%) and null (10%) severity levels to the error.
- 4th sentence: In overall, 95% of experts identified errors present in the sentence. 70% of them marked the **accuracy error** by noting the segment *rekomendavo jo naudoti* (Eng. recommended that it should not be used), whilst 30% of experts marked the **accuracy error** by noting the segment *į skleidimą* (Eng. the rollout). Collectively, 35% of experts marked the error as critical, 60% as major and 5% as minor.
- 5th sentence: 100% of experts identified the **accuracy error** by noting the word *jab* (Eng. jab) and assigned critical (70%), major (25%) and minor (5%) severity levels to it.
- 6th sentence: 80% of experts identified the **fluency issue** by noting the segment *apsaugo nuo sunkių ligų ši vakcina* (Eng. are protected from severe disease by this vaccine) and assigning the critical (5%), major (32%), minor (58%) and null (5%) severity levels to it. The severity of the untranslated word left the fluency error of unclear reference in the word *tai* (Eng. it) unnoticed.
- 7th sentence: 90% of experts identified errors present in the sentence. 70% of them marked the **accuracy error** by noting the word *Profas* (Eng. Prof), whilst 30% of experts marked the **accuracy error** by noting the word *reguliuotojai* (Eng. regulators). Collectively, 65% of experts marked the error as major, 25% as minor and 10% as null.

The results of the survey presented to experts show that experts identified errors in every sentence produced by Google Neural Machine Translation System and indicated that they would correct each and every sentence. The experts indicated 7 accuracy errors and 2 fluency errors in total. 6 of the indicated accuracy issues fall under the second-level subtype “Mistranslation”, and the remaining one falls under the second-level subtype “Untranslated”, whilst the indicated fluency issues fall under the second-level subtypes “Spelling” and “Grammar”. It must first be noted that the results of the experts’ survey showed that accuracy and fluency issues were viewed as more disturbing than locale convention or style issues as the latter were unnoticed by the experts. For example:

18. *Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.*
19. *Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.*

The overly literal translation of the segment in the 2nd sentence (18th example above) “to have the vaccine live on TV” – *vakciną tiesiogiai transliuoti per televiziją* – was noted by 100% of the experts.

However, the experts did not identify the locale convention issue concerned with the format of the name *Carstenas Watzlas*. Even though the Lithuanian alphabet does not include the “w” letter and the name was not adapted to the Lithuanian locale phonetically (“Carsten” is translated as *Carstenas* with “C” and not with “K”; “Watzl” is translated as *Watzlas* with “z” and not with “s” (and with the already mention “W” letter), this lack of adaptation did not catch the attention of any of the experts. The same situation is seen in the 7th sentence (see 19th example above); the lack of name adaptation was overlooked, the overly literal style of the segment *pakeis kursą* was unnoticed, and experts only indicated the accuracy issue in the word *Profas*. The reasons of such overlook are ambiguous; either locale convention issues and style do not have any influence over the acceptability of a text, or the issues of accuracy present in the same sentence disrupt it to an extent that they are viewed as the core and the most important issues. The latter reason is vindicated by the severity levels that experts assigned to the accuracy errors. As it is seen in Table 2, experts assigned critical and major severity levels to issues found in the 1st, 2nd, 4th, 5th and 7th sentences which contain mistranslation errors (i.e., second-level subtype of accuracy issues; see p. 34 and Appendix 4). Also, the severity levels that were assigned when assessing machine translation quality (see section 2.3 and Appendix 4) showed that locale convention and style issues are either minor or null and that their weights are 1 or 0 accordingly. These results supplement one another and show that accuracy issues are identified easier than locale convention or style issues due to the impact – the bigger weight – that they have on the quality as well as acceptability of the translation than the latter.

Another important aspect to note is that the experts found certain fluency issues to be more disturbing than others that is firstly justified by the overlook of fluency issue that seen in the following examples:

20. *JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jab.*

21. „*Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.*“

100% of experts identified the second-level subtype – “Untranslated” – accuracy issue in the word *jab* (Eng. jab) present in the 5th sentence (see 20th example), and 80% of experts identified the second-level subtype – “Grammar” – fluency issue in the segment *apsaugo nuo sunkių ligų ši vakcina* (Eng. are protected from severe disease by this vaccine) in the 6th sentence of the short excerpt (see 21st example). However, neither of the experts noted the second-level subtype – “Ambiguity – fluency issue present in the word *tai* (Eng. it) (see 21st example). Even though one may guess what *tai* refers to, unanimous views that experts share on the critical error in the word *jab* makes this reference illogical and unclear. However, the results showed that experts were more concentrated towards the fluency issue – “Grammar” – in the segment *apsaugo nuo sunkių ligų ši vakcina* and did not pay attention to the different fluency error present in the same sentence. Even though 58% of the experts marked the grammar error as minor, neither of them identified the unclear reference error present in the word *tai* (Eng. it) which was directly linked to the critical accuracy issue present in the word *jab* (Eng. jab). This shows that grammar errors were found to be more disturbing than even those errors of ambiguity that are directly linked to critical accuracy issues as grammar errors disrupt the overall flow of the sentence even though they are indicated as minor by experts. It also signals that the experts viewed and evaluated the sentences individually and not as a part of a text.

Moreover, it became apparent that certain accuracy issues catch more attention of experts than others. For instance:

22. „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.

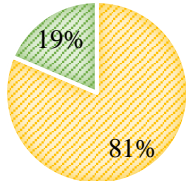
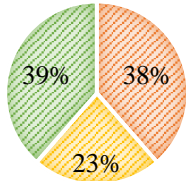
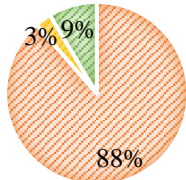
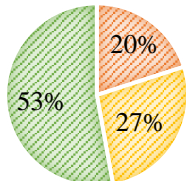
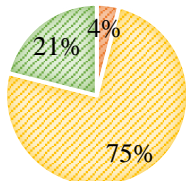
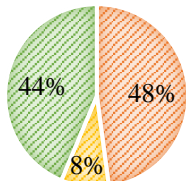
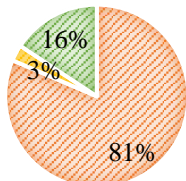
100% of experts identified the accuracy issue in the word *žandikaulį* (Eng. jab); however, neither of them noted the word *paremti* (Eng. to back) that was also marked as a second-level accuracy issue – “Overly literal translation” – in the translation quality assessment (see section 2.3 and Appendix 4). This shows that an error of the critical severity level obtained more attention than an error of the minor severity level (see Appendix 4). The overlook of the accuracy issue in some way resembles the overlook of the fluency issue as seen in the 6th sentence (see discussion of the 21st example on p. 42) in which a more severe error was indicated by experts. It thus became apparent that experts’ acceptability is mostly impacted by the more severe errors present in a sentence; therefore, it could be stated that the experts’ acceptability of the machine translation is impacted by the errors that have more weight on the overall translation quality score. This is also vindicated by the identification of the accuracy error in the word *žandikaulį* (Eng. jab) but the overlook of the accuracy error in the word *paremti* (Eng. to back) that showed that not only accuracy issues cause more comprehension issues than fluency, style or locale convention issues, but also that critical accuracy issues have the most impact on the quality and the acceptability of the machine translation as perceived by the experts.

Taking everything into account, the results of the experts’ survey analysis showed that the acceptability of the machine translation is the most affected by critical accuracy errors. The experts noted the minor severity level of spelling issues and did not draw attention to the unidiomatic style of certain sentences as well as the lack of adaptation of names to the Lithuanian locale which showed that experts – i.e., individuals who have knowledge and experience in translation – do not view style, locale convention or certain fluency issues to be the cause of a low translation quality and poor acceptability of the machine translation. It was also vindicated by the fact that style and locale convention issues obtained less or no attention of the experts if the same sentence contained accuracy issues. It must also be noted that experts unanimously identified grammar and spelling errors; however, they agreed that such fluency issues do not have a significant impact on the translation quality as they were mostly rated as minor. On the contrary, accuracy issues were rated as either critical or major whilst the indicated locale convention and styles issues were overlooked so they received null error weights. The severity levels assigned by the experts also mostly overlapped with the severity levels assigned in the translation quality assessment (see Appendix 4); however, it was noticed that experts shared far more critical views on the severity of the errors. These critical views also correspond to the experts’ opinion on whether the found errors should be corrected as all of them indicated that they would correct the errors. Furthermore, it was found that certain accuracy errors have more impact on the quality as well as acceptability of the machine translation as perceived by the experts than others as certain accuracy errors were left unnoticed if the same sentence contained another accuracy error of a higher severity level. Therefore, the analysis showed that the quality of the machine translation – therefore, the acceptability of it – is mostly influence by the critical accuracy errors present in it as perceived by the experts. Having analysed the results of the experts’ survey on the acceptability of the machine translation, the analysis of the non-experts’ survey results is expected to provide different results due to the more various backgrounds of the participants.

2.4.2. Relationship between errors and non-experts’ acceptability of machine translation

The same excerpt from the analysed machine translation presented to non-expert participants of the study has shown more various results. These are presented in the Table 3.

Table 3. Non-experts' evaluation of machine translation

–	Translated sentence	Answers to the question “Is the sentence comprehensible?”	Incomprehensible segment indicated by a non-expert
1	„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	66 answers (94% of non-experts) – <i>žandikaulį</i>
2	Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	41 answers (58% of non-experts) – <i>vakciną tiesiogiai transliuoti</i>
3	ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakciną naudojimą visoms amžiaus grupėms.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	6 answers (8% of non-experts) – <i>mėn. Patvirtino</i> 1 answer (2% of non-experts) – <i>ES medicinos priežiūros institucija</i>
4	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	34 answers (48% of non-experts) – <i>rekomendavo jo naudoti</i> 13 answers (18% of non-experts) – <i>reagavo į skleidimą</i>
5	JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jаб.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	64 answers (92% of non-experts) – <i>jab</i>
6	„Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.“	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	22 answers (32% of non-experts) – <i>apsaugo nuo sunkių ligų ši vakcina</i> 6 answers (8% of non-experts) – <i>sukelia imuninį atsaką</i>
7	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.	 <ul style="list-style-type: none"> ■ Yes ■ No ■ Partially 	12 answers (17% of non-experts) – <i>Profas</i>

- 1st sentence: 81% of non-experts identified the sentence as incomprehensible and 19% as partially comprehensible. 94% of non-experts marked the **accuracy issue** by noting the word *žandikaulį* (Eng. jab).
- 2nd sentence: The results showed a great variety of answers. 38% of non-experts found the sentence to be comprehensible, 39% – partially comprehensible, and 23% – incomprehensible. 58% of non-experts identified the **accuracy issue** by noting the segment *vakciną tiesiogiai transliuoti* (Eng. to have the vaccine live on). However, just as the experts, non-experts did not identify the locale convention issue present in the name *Carstenas Watzlas* (Eng. Carsten Watzl).
- 3rd sentence: The results showed that the sentence caused the least comprehension issues to non-experts as 88% of them found the sentence to be comprehensible (9% – partially comprehensible, 3% – incomprehensible). Only 8% of non-experts identified the **fluency issue** by noting the segment *mėn. Patvirtino* (Eng. approved the use of ... in January), whilst 2% also found issues in the segment *ES medicinos priežiūros institucija* (Eng. The EU's medical regulator) which was not identified as an error in the translation quality analysis (see section 2.3 and Appendix 4).
- 4th sentence: The sentence was viewed differently by non-experts. 20% of them found the sentence to be comprehensible, 53% – partially comprehensible, 27% – incomprehensible. The great variation of the results arises because of the indicated errors; 48% of non-experts identified the **accuracy issue** by noting the segment *rekomendavo jo naudoti* (Eng. recommended that it should not be used), whilst 18% of non-experts identified a different **accuracy issue** by noting the segment *reagavo į skleidimą* (Eng. But the rollout was met).
- 5th sentence: The results showed the unity of non-experts as 75% of them found the sentence to be incomprehensible (21% – partially comprehensible, 4% – comprehensible). 92% of non-experts identified the reason of the low comprehensibility of the sentence, which was the **accuracy issue** in the word *jab* (Eng. jab).
- 6th sentence: The views on the comprehensibility of the sentence differed. 48% of non-experts found the sentence to be comprehensible, 44% – partially incomprehensible, and 8% – incomprehensible. 32% of non-experts identified the **fluency issue** by noting the segment *apsaugo nuo sunkių ligų ši vakcina* (Eng. are protected from severe disease by this vaccine), and 8% of them also found issues in the segment *sukelia imuninį atsaką* (Eng. elicits an immune response) which was not identified as an error in the translation quality analysis (see section 2.3 and Appendix 4).
- 7th sentence: The results showed that non-experts found it mostly comprehensible (81% and 16% – partially comprehensible), whilst only 3% found it to be incomprehensible. 17% of non-experts identified the reason of the incomprehensibility or the partial comprehensibility of the sentence by noting the **accuracy issue** in the word *Profas* (Eng. Prof).

The results of the survey presented to non-experts showed that non-experts were no less tolerant towards the errors present in machine translation than experts. The data in Table 3 presents more optimistic views towards the comprehensibility – thus, acceptability – of the machine translation than the views shared by experts (see Table 2) in a way that the majority of the sentence were marked as comprehensible or at least partially comprehensible by non-experts. Despite that, non-experts took the task seriously and indicated even more errors in the sentences than experts did. In total, non-experts marked 10 errors, 6 of which were accuracy issues and 2 fluency issues. 5 of the indicated

accuracy issues fall under the second-level subtype “Mistranslation”, and the remaining one falls under the second-level subtype “Untranslated”, whilst the indicated fluency issues fall under the second-level subtypes “Spelling” and “Grammar”.

Non-experts shared the most consensus on the 1st and 5th sentences indicated below:

23. „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.

24. JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jab.

81% of non-experts identified the 1st sentence (see 23rd example above) as incomprehensible, whereas 75% of them also identified the 5th sentence (see 24th example above) as incomprehensible. The significantly low comprehensibility of the sentences is vindicated by the segments identified by non-experts, i.e., 94% of non-experts noted the second-level accuracy issue – “Mistranslation” – in the word *žandikaulį* (Eng. jab) in the 1st sentence, and 92% of non-experts noted the second-level accuracy issue – “Untranslated” – in the word *jab* (Eng. jab) in the 5th sentence. These errors were identified as critical in the translation quality assessment part (see Appendix 4) for the significant impact they have on the target text and the ideas present in it; the weight and the impact of the critical accuracy errors corresponds with the non-experts’ views on the comprehensibility of the sentences containing such errors.

The other accuracy error present in the 1st sentence (see example 23) was overlook by non-experts. The second-level accuracy error – “Mistranslation” – in the overly literally translated word *paremti* (Eng. to back) was unnoticed by non-experts. This might have occurred due to two reasons: either the overly literal translation of the word did not have negative impact on the comprehensibility of the translation (which is also vindicated by the minor severity level of the error (see Appendix 4), or the critical accuracy error present in the sentence (i.e., *žandikaulį* (Eng. jab) diminishes the importance and the significance on the other accuracy issues present in the same sentence. The latter is also justified by the results of the experts’ survey in which the respondents also overlooked the minor accuracy error and indicated only the critical one.

Other issues that were noted by non-experts as incomprehensible were found in the following sentences:

25. Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.

26. Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.

58% of non-experts found the overly literal translation in the segment *vakciną tiesiogiai transliuoti* (Eng. to have the vaccine live on TV) even though only 23% of them marked the whole sentence (see 25th example above) as incomprehensible. The bigger number of respondents who indicated the incomprehensible segment than the number of respondents who marked the sentence as incomprehensible shows that even though the accuracy error present in the sentence does disturb the reading pattern and does have influence on the comprehensibility of the translation, it is not viewed by non-experts as critical which coincides with the translation quality assessment (see Appendix 4) in which the overly literal translation of the segment was marked as a major error that requires more attention of a reader, but does not have a significant impact on the overall idea of the text.

Furthermore, an important aspect to note that is neither of the 70 non-experts who participated in the experiment noted the error in the word *Carstenas Watzlas* (Eng. Carsten Watzl) which not only included the “w” letter which is not present in the Lithuanian alphabet, but also was not adapted phonetically. This also corresponds with the results of the experts’ survey in which neither of the respondents marked the locale convention issue present in the sentence. This allows to draw firm conclusions that accuracy errors cause more comprehension issues than locale convention errors.

Contrary to the 2nd sentence (see 25th example above), critical accuracy errors trigger different reactions on the comprehensibility of a text as perceived in the 4th sentence (see 26th example above). 53% of non-experts marked the sentence as partially comprehensible, 27% as incomprehensible, and only 20% as comprehensible due to the two errors found in the sentence. The first second-level accuracy error – “Mistranslation” – was identified in the segment *rekomendavo jo naudoti* (Eng. recommended that it should not be used) by 48% of non-experts, and 18% of non-experts also identified the second-level accuracy error – “Mistranslation” – in the segment *reagavo į skleidimą* (Eng. But the rollout was met). The bigger number of non-experts who identified the former error shows the bigger significance of the error which also corresponds with the translation quality assessment results (see Appendix 4) which showed that the first accuracy error present in the sentence is of a critical severity level, whereas the second is of a major severity level. This once again proves that critical accuracy errors have the biggest impact on the machine translation quality and non-experts’ acceptability of it.

One other sentence that received considerably ambiguous results on the comprehensibility as deemed by non-experts was the following:

27. „*Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.*“

The sentence was deemed considerably comprehensible as 48% of non-experts thought the sentence to be comprehensible, and 44% – partially comprehensible, whilst only 8% marked the sentence as incomprehensible. Despite the high score, 40% non-experts in total identified issues in the sentence; 32% marked the second-level fluency issue – grammar – in the segment *apsaugo nuo sunkių ligų ši vakcina* (Eng. are protected from severe disease by this vaccine), and 8% marked the segment *sukelia imuninį atsaką* (Eng. it elicits an immune response). Similarly to the previously discussed 2nd sentence (see 25th example), a considerably high number of non-experts identified an incomprehensible segment in a sentence that was marked as mainly comprehensible or partially comprehensible; this once again indicates that the error identified by non-experts has influence on the machine translation quality, but the weight of it is not as significant as to lower the non-experts’ acceptability of the machine translation. Moreover, non-experts did not take into account the fluency issue in the word *tai* (Eng. it), in which a second-level issue – “Unclear reference” – was identified. This once again provides ambiguous results as it did in the analysis of the experts’ survey results (see section 2.4.1). Also, an interesting aspect to note is that even though non-experts marked identified the fluency issue, they also noted that the segment *sukelia imuninį atsaką* (Eng. it elicits an immune response) contains issues. However, no error was identified in the segment after carrying out the machine translation quality analysis (see Appendix 4); therefore, it could be assumed that the respondents who identified the issue in the segment are not familiar with the very basic medical terminology. A similar example was also provided in the results of the 3rd sentence:

28. *ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinų naudojimą visoms amžiaus grupėms.*

2% of non-experts marked the segment *ES medicinos priežiūros insitucija* (Eng. The EU's medical regulator) as incomprehensible; however, just like in the segment analysed above, no error was identified after carrying out the machine translation quality analysis (see Appendix 4). For this reason, it can once again be concluded that 2% of non-experts who marked the segment as incomprehensible did so because of lack of knowledge in the specific terminology. Nevertheless, the sentence was viewed as comprehensible by 88% of non-experts, partially comprehensible by 9%, and incomprehensible only by 3%. 8% identified the reason of the incomprehensibility to be the segment *mėn. Patvirtino* (Eng. approved ... in January), in which a second-level fluency issue – “Spelling” – was identified. However, the low number of respondents who deemed the sentence incomprehensible and the minor severity level of the error signifies that the fluency issue does not have a significant influence on the acceptability of the machine translation as perceived by non-experts.

The last sentence of the short excerpt was also deemed as considerably comprehensible by non-experts as 81% of them marked the sentence as comprehensible, 16% as partly comprehensible, and 3% and incomprehensible:

29. *Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.*

32% of non-experts noting the word *Profas* (Eng. Prof) as incomprehensible which was identified as a second-level accuracy issue – “Mistranslation” – when assessing machine translation quality (Appendix 4). The same sentence also contained the style issue in the segment *pakeis kursą* (Eng. reverse course) that was translated literally this way losing the idiomaticity of the source text. However, the unconveyed style had no weight over the translation quality and it did not have any impact on the acceptability of the machine translation as perceived by non-experts.

Taking everything into account, the results of the non-experts' survey analysis showed that even though it was expected that non-experts would share more positive views towards the quality of the machine translation due to their non-linguistic backgrounds or lack of experience in translation, the acceptability of the machine translation was affected equally negatively by the errors present in the translation as for the experts. It must be noted that non-experts did not note the significance of style or locale convention issues as they neither identified such errors nor marked the segments containing style or locale convention issues as incomprehensible. The non-experts also shared a unanimous opinion towards the fluency issues and did not mark the significant importance and impact of spelling errors on the comprehensibility of the machine translation; however, they agreed that second-level fluency issues of grammar disrupt the flow of a sentence and introduce unnaturalness to it, which is one of the aspects that make a translation unacceptable in terms of its quality. Moreover, it was found that non-experts were the most disturbed by accuracy issues as second-level errors of mistranslated or untranslated units were the main cause of incomprehensibility – thus, low acceptability of the machine translation – as the sentences that contained untranslated or mistranslated units obtained the lowest comprehensibility scores as rated by non-experts. Also, an interesting result to note is that the presence of accuracy issues in a sentence reduced the significance of other issues (e.g., style, fluency or locale convention issues) as non-experts overlooked other issues if mistranslated or untranslated units were included in a sentence. In addition, a similar sequence was noticed if a sentence contained

more than one accuracy issue – one of them was most often overlooked by non-experts. The analysis showed that accuracy errors that were identified as critical and had more weight in the translation quality assessment were noted more easily by non-experts, and the errors that were identified as major, minor or null were overlooked if the same sentence contained critical accuracy errors. Conversely, if a sentence contained several accuracy errors of the same severity, they both were identified by non-experts. Therefore, it could be said that the non-experts were disturbed by fluency issues; however, the comprehensibility, thus acceptability of the machine translation was the most influenced by the critical accuracy errors present in the translation.

To conclude the third analytical part of the work, the analysis of experts and non-experts' surveys on the acceptability of the machine translation showed that non-experts share considerably similar views to those of the experts. The similarity is apparent in the number of the indicated issues as well as the types of them and the highlighted significance of accuracy issues. The results demonstrate that both experts and non-experts were not affected by locale convention and style issues and that both groups identified fluency and accuracy issues present in the translation as the cause of incomprehensibility of the text; most importantly, all of the respondents also marked critical accuracy errors to be the main cause of the low quality of the machine translation. Also, even though non-experts indicated more incomprehensible segments than experts did, the results of the surveys show that non-experts found the machine translation more acceptable than the experts who identified errors in all of the sentences and indicated that all of them should be corrected. Therefore, it could be stated that non-experts found the machine translation to be more acceptable in terms of quality than experts did; however, all the respondents' views coincided on the fact that the machine translation quality is mostly influenced by the critical accuracy issues – mistranslated, untranslated or omitted units.

Having analysed the results of experts and non-experts' surveys on the acceptability of the machine translation and having assessed the machine translation quality, it became apparent that both experts and non-experts do not view the translation as acceptable mainly due to the low quality of it. The unacceptability is first vindicated by the results obtained by employing the one-way MANOVA which showed that neither of the six presented statements achieved the mean value of 3 or higher, meaning that most of the answers to all the statements concerning the three elements of acceptability – satisfaction, quality, usability – were somewhat disagreed or completely disagreed with. Nonetheless, the standard deviations showed that experts and non-experts slightly disagree on satisfaction with and usability of the machine translation in a way that non-experts shared considerably more satisfaction with the way the main idea is rendered than experts who disagreed with the statement, and in a way that experts unanimously disagreed that the text is useful and informative whereas non-experts shared various views but agreed with the experts that the source is unreliable. In overall, non-experts were more satisfied with the machine translation and found it more usable than experts; however, their views on the quality of the machine translation coincided. All of the respondents agreed that the translation quality is low and that the text sounds unnatural, which made the 3rd element of acceptability – quality – the least acceptable aspect of the machine translation.

The assessment of the machine translation quality in the second part of the analysis justified the low results obtained from the experts and non-experts' surveys on the acceptability. A big number of errors (80) of different severity levels found in the machine translation consisting of 532 words resulted in a negative translation score of -0.364 , which was impacted by accuracy, fluency, style and locale convention issues. The analysis showed that the machine translation contained 45 (56%) accuracy issues, 18 (22.5%) fluency issues, 3 (3.8%) style issues, and 14 (17.5%) locale convention

issues, and it was found that the penalties score of the accuracy issues was 96%, which indicated that such errors as mistranslated, untranslated or omitted units have the most impact on the overall translation quality, whilst spelling, grammar, unidiomatic style, lack of name or punctuation adaptation accounted for only 4% of the overall penalties' number, therefore, the overall translation quality score. It must also be noted that not all accuracy errors had such a significant impact on the overall quality of the translation. The translation quality assessment showed that whilst the majority of the found accuracy, fluency, style and locale convention errors were either of major or minor severity levels, the considerably few found critical accuracy errors impacted the quality of the translation the most.

The results of machine translation quality assessment corresponded with the results of the surveys. Critical accuracy errors were identified by experts as critical, and the sentences that contained these errors were identified by non-experts as incomprehensible in the third part of the analysis. An important aspect to note is that experts and non-experts' acceptability of the machine translation in terms of quality was the least impacted by locale convention and style issues as the presence of those was identified by neither of the groups. Moreover, both experts and non-experts marked that fluency issues have negative impact on the comprehensibility of the text; however, the significance of those is not great as the weight of grammar, spelling or ambiguity errors is relatively low as compared with the weight that critical accuracy issues such as mistranslation or omission has. The significance of the accuracy issues was also vindicated by the fact that neither experts nor non-experts indicated errors of style, fluency or locale convention if the same sentence contained accuracy errors. For this reason, it can be concluded that analysis showed that all of the surveys' respondents agree that accuracy errors are the main reason for the low quality of a machine translation, which corresponds to the results of the machine translation quality assessment and supplements the low acceptability scores. Nonetheless, it also should be noted that experts were more critical towards all the found mistakes in the machine translation as they marked the majority of the errors to be major or critical, whereas a significantly lower percentage of non-experts marked the segments containing the same errors as incomprehensible. This shows that experts are concentrated towards a faultless translation, whilst non-experts are not significantly disturbed by minor fluency, accuracy, style or locale convention issues or even major accuracy or fluency errors and may deem a translation acceptable despite the presence of those.

Thus, it could be concluded that machine translation is not yet deemed acceptable by basic users and translators as the expectations of satisfaction, usability and quality are not met. Whilst non-experts share more positive views towards satisfaction and usability of the machine translation, all of the respondents who participated in the surveys unanimously agree that the translation quality is low. The low translation quality score obtained after the assessment of the machine translation showed that the quality is mostly affected by critical accuracy errors that distort the idea of the text which was also agreed by experts and non-experts as all the respondents marked the incomprehensibility of the segments containing such errors. Therefore, it could be said that critical accuracy errors are the main reason of the low machine translation quality as well as low acceptability of the machine translation output and that the anticipation of such errors prior to publishing such content to public media channels could prove to deliver more reaction as well as more response to the ideas present in a text.

Conclusions

1. The reason of a number of different acceptability definitions in translation is the subjectivity of the notion. Human acceptability of machine translation is understood as a text user's subjective opinion on the target text; for a translation to be deemed acceptable, it has to meet the needs as well as expectations of its user and to contain no significant errors that would disrupt the presented idea or the flow of the text. It has, therefore, been suggested that acceptability should be measured by considering three elements: satisfaction, usability, and quality. Also, considering the subjectivity of the matter, acceptability of machine translation can only be measured by its users, preferably by providing questionnaires or online surveys. The latter is especially recommended as only a large number of answers would allow drawing conclusions on the subjective notion of acceptability. Moreover, it is apparent that acceptability is always influenced by errors present in a translation; for this reason, attention should be drawn to the types of issues present in a machine translation which also allow identifying the weakest aspects of modern translation systems.
2. A great variety of different error taxonomies have been introduced to assess machine translation quality. However, this big number of different methodologies does not eliminate the subjectivity and complexity of the matter as results on machine translation quality may still vary depending on the errors that the different taxonomies distinguish. For this reason, there is a need of a universal translation quality assessment system which could be applied in the evaluation of a translation product for any language direction. Several attempts have been made to create standardised translation quality assessment systems, one of which is a Multidimensional Quality Metrics introduced in EU funded QTLaunchPad project. The metrics are tailored in a way that they can be used for the evaluation of both human and machine translations. MQM adopts a functionalist approach that allows analysing only those issue dimensions that are relevant for the purposes of an individual research. Also, the strict hierarchy of MQM presents 10 dimensions under which 182 issue types are presented, and clear definitions are given for each and every issue type. The clear issue definitions and the functionalist approach of MQM makes them an appealing and one of the most viable metrics to be used in machine translation quality assessment.
3. Machine translation of a general text taken from a news website is deemed unacceptable by both experts and non-experts. It is found that neither of the three elements of acceptability are met as all respondents agree that the machine translation does not meet their needs and expectations. However, it should be noted that non-experts share more positive views towards satisfaction and usability of machine translation whilst experts unanimously agree that the translation can be deemed neither satisfactory nor usable. The opinion of all respondents coincides in the evaluation of quality as neither of them agreed that the quality of machine translation is high. The results confirm the first hypothesis that machine translation from English to Lithuanian is deemed more acceptable by non-experts than by experts. Also, the results show that the low quality of machine translation is the main factor that makes machine translation unacceptable, which confirms the second hypothesis that quality issues have the most influence on the machine translation acceptability.
4. The machine translation quality analysis results coincide with the results of acceptability as perceived by experts and non-experts as the obtained translation quality score is low due to a considerably big number of the found issues and their severity levels. The quality is influenced by accuracy, fluency, style and locale convention errors that cause comprehension issues. It is found that most of the issues present in the translation are related to accuracy dimension, followed by issues of fluency, locale convention and style. Furthermore, machine translation quality

assessment shows that machine translation is not yet tailored towards the Lithuanian locale that is vindicated by the usage of locale-unspecific punctuation or lack of name format adaptation. The unnaturalness of machine translation output is also caused by grammar, spelling and typography errors; however, the considerably low severity levels of both locale convention and fluency errors do not have a significant impact on the machine translation quality as they do not disrupt the content to an extent that would make it unfit for use. The quality is mostly lowered by the critical accuracy errors that present ambiguousness and inaccuracy of the conveyed information in a way that the meaning is changed completely. These errors have the most influence on the overall machine translation quality score, and the rendering of the inaccurate information is deemed the weakest aspect of the modern machine translation system.

5. The results show that even though the backgrounds of the respondents differed, the views on the factors that make a translation unacceptable in terms of quality coincided. The errors of style and locale convention were left unnoticed, which indicates that users of machine translation systems are not disturbed by the presence of foreign elements in a text as well as the lack of compliance with Lithuanian language norms in regard to the rendering of punctuation and foreign names. However, both experts and non-experts are found to be disturbed by the presence of fluency issues that are apparent in the incorrect word orders or unnecessary capitalisation, though neither of the groups indicate that fluency issues cause significant comprehension issues. On the contrary, accuracy issues are unanimously marked as causing the most comprehension issues. If a sentence contains issues of accuracy and any other dimension, only accuracy issue is identified; if a sentence contains several accuracy issues, only a more severe – critical – accuracy issue is identified. This shows that the presence of critical accuracy issues is the main factor that negatively influences the overall machine translation quality – thus acceptability – whilst fluency issues are identified as disturbing yet not major, and issues of style and locale convention are not perceived as having any influence over the comprehensibility of machine translation by both experts and non-experts.

List of references

1. Bader, M., & Häußler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, 46(2), 273–330.
2. Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude Estimation of Linguistic Acceptability. *Language*, 72(1), 32–68.
3. Bojar, O., Ercegovčević, M., Popel, M., & Zaidan, O. F. (2011). A grain of salt for the WMT manual evaluation. In *Proceedings of the 6th Workshop on Statistical Machine Translation, 30-31 July 2011, Edinburgh, Scotland* (pp. 1–11). Stroudsburg, Pennsylvania: Association for Computational Linguistics.
4. Byrne, J. (2006). *Technical Translation. Usability strategies for translating technical documentation*. Dordrecht: Springer.
5. Castilho, S. (2016). *Measuring acceptability of machine translated enterprise content*. Dublin: Dublin City University.
6. Castilho, S., & O'Brien, S. (2016). Evaluating the Impact of Light Post-Editing on Usability. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 23–28 May 2016, Portorož, Slovenia* (pp. 310–316). Luxembourg: European Language Resources Association (ELRA).
7. Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to Human and Machine Translation Quality Assessment. In Moorkens, Castilho, Gaspari, Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 9–38). Cham: Springer International Publishing.
8. Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., & Georgakopoulou, P. (2017). Evaluating MT for massive open online courses: A multifaceted comparison between PBSMT and NMT systems. *Machine Translation*, 32(1), 255–278.
9. Castillo, L. M. (2015). Acquisition of translation competence and translation acceptability: an experimental study. *Translation & Interpreting* 7(1), 72–85.
10. Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
11. Coughlin, D. (2003). Correlating Automated and Human Assessments of Machine Translation Quality. In *Proceedings of the Machine Translation Summit IX, 23–27 September 2003, New Orleans, USA*, (pp. 63–70). Stroudsburg, Pennsylvania: Association for Machine Translation in the Americas.
12. Daubariénė, A., & Ziezytė, G. (2013). Machine Translation: Translated Texts in Terms of Standards of Textuality. *Studies About Languages*, 22, 55–61.
13. De Beaugrande, R., & W. Dressier (1981). *Introduction to Text Linguistics*. New York: Longman.
14. Doherty, S., & O'Brien, S. (2014). Assessing the Usability of Raw Machine Translated Output: A User-Centered Study Using Eye Tracking. *International Journal of Human-Computer Interaction*, 30(1), 40–51.
15. Drugan, J. (2013). *Quality in Professional Translation: Assessment and Improvement*. London: Bloomsbury Academic.
16. Farrús, M., Costa-Jussà, M. R., Mariño, J. B., & Fonollosa, J. A. R. (2012). *Linguistic-based evaluation criteria to identify statistical machine translation errors* [accessed 2021-03-28]. Retrieved from https://www.researchgate.net/publication/44181697_Linguistic-based_evaluation_criteria_to_identify_statistical_machine_translation_errors

17. Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: WH-constraints in German. *Lingua*, 115(11), 1525–1550.
18. Federico, M., Negri, M., Bentivogli, L., & Turchi, M. (2014) Assessing the impact of translation errors on machine translation quality with mixed-effects models. In *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing, October 2014, Doha, Qatar* (pp. 1643–1653). Stroudsburg, Pennsylvania: Association for Computational Linguistics.
19. 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* [accessed 2021-05-05]. Retrieved from <https://arxiv.org/pdf/2104.14478.pdf>
20. Galesic, M. (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics*, 22(2), 313–328.
21. Gondra, A. (2018). Preverbal Double Negation in Basque Spanish? Acceptability Judgment Task and Semistructured Interview. *Hispania* 101(4), 587–604.
22. International Organization for Standardization (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability* (ISO Standard No. 9241-11:1998). Retrieved from <https://www.iso.org/standard/16883.html>
23. International Organization for Standardization (2002). *Ergonomics of human-system interaction – Usability methods supporting human-centred design* (ISO Standard No. 16982:2002). Retrieved from <https://www.iso.org/standard/31176.html>
24. Kasperavičienė, R., Motiejūnienė, J., & Patašienė, I. (2020). Quality Assessment of Machine Translation Output: Cognitive Evaluation Approach in an Eye Tracking Experiment. In *Texto livre: linguagem e tecnologia. Belo Horizonte* 13(2), 271–285.
25. Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects degrees of grammaticality*. Edinburgh: University of Edinburgh.
26. Kertz, L. (2013). Verb Phrase Ellipsis: The View from Information Structure. *Language*, 89(3), 390–428.
27. Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108, 121–132.
28. Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2018). Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian. *Machine Translation* 32, 195–215.
29. Koby, G. S., Fields, P., Hague, D., Lommel, A., & Meldy, A. (2014). Defining Translation Quality. *Tradumàtica: Tecnologies de La Traducció*, 12, 413–420.
30. Lassen, I. (2003). *Accessibility and Acceptability in Technical Manuals: A survey of style and grammatical metaphor*. Amsterdam: John Benjamins.
31. Lommel, A. (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In Moorkens, Castilho, Gaspari, Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109–127). Cham: Springer International Publishing.
32. Lommel, A., Burchardt, A., & Uszkoreit, H. (2015b). *Multidimensional Quality Metrics (MQM) Definition* [accessed 2021-04-15]. Retrieved from: <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html#introduction>

33. Lommel, A., Burchardt, A., Görög, A., Uszkoreit, H., & Melby, A. K. (2015a). *Multidimensional Quality Metrics (MQM) Issue Types* [accessed 2021-04-01]. Retrieved from: <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>
34. Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica: tecnologies de la traducció*, 12, 455–463.
35. Neubert, A., & Shreve, G. M. (1992). *Translation as Text*. Kent: The Kent State University Press.
36. Nielsen, J. (1993). *Usability Engineering*. Amsterdam: Morgan Kaufmann.
37. O'Brien, S., Balling, L. W., Carl, M., Simard, M., & Specia, L. (2014). *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing.
38. Ortiz-Boix, C., & Matamala, A. (2017). Assessing the quality of post-edited wildlife documentaries. *Perspectives*, 25(4), 571–593.
39. PACTE. (2008). First results of a Translation Competence Experiment: 'Knowledge of Translation' and 'Efficacy of the Translation Process'. In J. Kearns (Ed.), *Translator and Interpreter Training. Issues, Methods and Debates* (pp. 104–126). London: Continuum.
40. PACTE. (2009). Results of the Validation of the PACTE Translation Competence Model: Acceptability and Decision Making. *Across Languages and Cultures*, 10(2), 207–230.
41. Popović, M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In Moorkens, Castilho, Gaspari, Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 129–158). Cham: Springer International Publishing.
42. Reips, U. D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243–256.
43. Roturier, J. (2006). *An investigation into the impact of controlled English rules on the comprehensibility, usefulness and acceptability of machine-translated technical documentation for French and German users*. Dublin: Dublin City University.
44. Rubin, J., & Chisnell, D. (2011). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Indianapolis: Wiley.
45. Sirena, D. (2004). *Mission impossible: improve quality, time and speed at the same time* [accessed 2021-03-02]. Retrieved from: <https://www.translationdirectory.com/article387.htm>
46. Snow, T. A. (2015). *Establishing the viability of the multidimensional quality metrics framework* [accessed 2021-03-02]. Retrieved from: <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=6592&context=etd>
47. Sprouse, J. (2011). A Test of the Cognitive Assumptions of Magnitude Estimation: Commutativity Does Not Hold for Acceptability Judgments. *Language*, 87(2), 274–288.
48. Sprouse, J., & Almeida, D. (2012). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics*, 48(3), 609–652.
49. Stymne, S., Danielsson, H., Bremin, S., Hu, H., Karlsson, J., Lillkull, A. P., & Wester, M. (2012). Eye Tracking as a Tool for Machine Translation Error Analysis. In Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (Eds.), *Proceedings of the eighth international conference on language resources and evaluation, 23–25 May 2012, Instabul, Turkey* (pp. 1121–1126).
50. Suojanen, T., Koskinen, K. & Tuominen, T. (2015). *User-Centered Translation*. Abingdon: Routledge.

51. Temnikova, I. (2010) Cognitive evaluation approach for a controlled language post-editing experiment. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17–23 May 2010, Valletta, Malta* (pp. 3485–3490).
52. Temnikova, I., Zaghouni, W., Vogel, S., & Habash, N (2016). Applying the Cognitive Machine Translation Evaluation Approach to Arabic. In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, 23–28 May 2016, Portoroz, Slovenia* (pp. 3644-3651).
53. Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
54. Uszkoreit, H., & Lommel, A. (2013). *Multidimensional Quality Metrics: A New Unified Paradigm for Human and Machine Translation Quality Assessment* [accessed 2021-04-01]. Retrieved from: <http://www.qt21.eu/launchpad/sites/default/files/MQM.pdf>
55. Van Slype, G. (1979). *Critical study of methods for evaluating the quality of machine translation*. Brussels: Buerau Marcel van Dijk.
56. Vardaro, J., Schaeffer, M., & Hansen-Schirra, S. (2019). Translation Quality and Error Recognition in Professional Neural Machine Translation Post-Editing. *Informatics* 6(3), 1–29.
57. Venuti, L. (2008). *The Translator's Invisibility: A History of Translation* (2nd ed.). New York: Routledge.
58. Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error Analysis of Machine Translation Output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 697–702 [accessed 2021-03-04]. Retrieved from: https://www.researchgate.net/publication/307174612_Error_Analysis_of_Machine_Translation_Output
59. Way, A. (2018). Quality Expectations of Machine Translation. In Moorkens, Castilho, Gaspari, Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 159–178). Cham: Springer International Publishing.
60. Weskott, T., & Fanselow, G. (2011). On the Informativity of Different Measures of Linguistic Acceptability. *Language*, 87(2), 249–273.
61. Williams, M. (2004). *Translation Quality Assessment: An Argument-Centered Approach*. Ottawa: University of Ottawa Press.
62. Williams, M. (2009). Translation Quality Assessment. *Mutatis Mutandis*, 2(1), 3–23.
63. Ye, Y., & Toral, A. (2020). Fine-grained Human Evaluation of Transformer and Recurrent Approaches to Neural Machine Translation for English-to-Chinese. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation. 3–5 November 2020, Lisboa, Portugal* (pp. 125–134). European Association for Machine Translation.
64. Zamanian, M., & Heydari, P. (2012). Readability of Texts: State of the Art. *Theory and Practice in Language Studies*, 2(1), 43–53.

List of information sources

1. News website “BBC”. Online access: <https://www.bbc.com/>

Appendices

Appendix 1. Source text and target text

Source text (retrieved from https://www.bbc.com/news/world-europe-56227171)	Target text
<p>Covid vaccine: Germany urged to back AstraZeneca jab for over-65s</p> <p>A senior German immunologist has urged his country to change its mind and start allowing over-65s to receive the Oxford-AstraZeneca vaccine.</p> <p>In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.</p> <p>Carsten Watzl urged Angela Merkel to have the vaccine live on TV to prove it is safe.</p> <p>Germany's vaccine commission is currently reviewing its recommendation.</p> <p>Prof Watzl's call comes after recent studies in Scotland showed the AstraZeneca jab to be effective among the elderly.</p> <p>Germany is one of several EU states that have expressed doubts over the efficacy of the vaccine in older people.</p> <p>The country is currently struggling to avoid a third wave of infections as cases remain stubbornly high.</p> <p>The EU's medical regulator approved the use of the Oxford-AstraZeneca vaccine for all age groups in January.</p> <p>But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65. They cited insufficient data on its efficacy for older people.</p> <p>German health authorities have so far used fewer than 300,000 of the 1.17 million doses of the AstraZeneca vaccine the country has received.</p> <p>In January, French President Emmanuel Macron said the jab was "quasi-ineffective" for older age groups - a claim strongly refuted by the UK government and British medical regulators. AstraZeneca itself says the vaccine is effective at all ages.</p> <p>The UK is among countries that have approved the jab for all age groups.</p> <p>The decision was boosted by recent research led by Public Health Scotland, which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.</p> <p>Among the over 80s, there was an overall 81% reduction in the numbers admitted to hospital when the results for both vaccines were combined.</p> <p>"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.</p>	<p>„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms</p> <p>Vyresnysis vokiečių imunologas paragino savo šalį persigalvoti ir leisti leisti vyresniems nei 65 metų asmenims skiepytis „Oxford-AstraZeneca“.</p> <p>BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių.</p> <p>Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.</p> <p>Vokietijos vakcinų komisija šiuo metu peržiūri jos rekomendaciją.</p> <p>Prof. Watzlio kvietimas įvyko po to, kai neseniai Škotijoje atlikti tyrimai parodė, kad „AstraZeneca“ žandikaulis yra veiksmingas vyresnio amžiaus žmonėms.</p> <p>Vokietija yra viena iš kelių ES valstybių, kurios išreiškė abejones dėl vakcinų veiksmingumo vyresnio amžiaus žmonėms.</p> <p>Šiuo metu šalis stengiasi išvengti trečios infekcijų bangos, nes atvejų vis dar atkakliai daug.</p> <p>ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinų naudojimą visoms amžiaus grupėms.</p> <p>Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms. Jie nurodė nepakankamus duomenis apie jo veiksmingumą vyresnio amžiaus žmonėms.</p> <p>Vokietijos sveikatos priežiūros institucijos kol kas sunaudavo mažiau nei 300 000 iš šalyje gautų 1,17 milijonų vakcinų „AstraZeneca“ dozių.</p> <p>Sausio mėn. Prancūzijos prezidentas Emmanuelis Macronas teigė, kad vyresnio amžiaus žmonėms šiukšlės buvo „beveik neveiksmingos“ - JK vyriausybė ir Didžiosios Britanijos medicinos priežiūros institucijos tvirtai paneigė šį teiginį. Pati „AstraZeneca“ teigia, kad vakcina yra veiksminga bet kokio amžiaus.</p> <p>JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jabs.</p> <p>Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.</p> <p>Tarp vyresnių nei 80 metų pacientų, paguldytų į ligoninę, skaičius sumažėjo 81%, kai buvo susieti abiejų vakcinų rezultatai.</p>

<p>"In order for us to reach our vaccination goals we need people to get this vaccine.</p> <p>"We do know that the vaccine works in that age group. The recent data from Scotland clearly show it elicits an immune response, the elderly are protected from severe disease by this vaccine."</p> <p>Prof Watzl predicted German regulators would "reverse course based on the data coming out of Scotland".</p> <p>"If at that point [Chancellor] Angela Merkel were to go on live television and have the vaccine, that would be great," he added.</p> <p>On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated recommendation" on the AstraZeneca vaccine.</p> <p>He said it was waiting for more details from the authors of the study carried out in Scotland.</p> <p>"We have never criticised the vaccine," he said, describing it as "very good".</p> <p>Germany has confirmed more than 2.4 million cases of the coronavirus since the pandemic began and just over 70,000 deaths.</p> <p>A lockdown has been in place since mid-December but infections are still worryingly high and officials have found it difficult to ease restrictions.</p> <p>Figures released last week showed the infection rate had risen to 59.3 cases per 100,000 people over the past seven days, from 57 a week earlier.</p> <p>The government is due to meet next week to consider lockdown rules that run to 7 March.</p> <p>Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.</p>	<p>"- Manau, kad Vokietija taip pat greitai pakeis kursą" - sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.</p> <p>„Kad galėtume pasiekti užsibrėžtus skiepavimo tikslus, reikia, kad žmonės gautų šią vakciną.</p> <p>"- Mes tikrai žinome, kad vakcina veikia toje amžiaus grupėje. Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina."</p> <p>Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.</p> <p>„Jei tuo metu [kanclerė] Angela Merkel eitų į tiesioginę televiziją ir turėtų vakciną, būtų puiku“, - pridūrė jis.</p> <p>Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą, AstraZeneca “vakcinos rekomendaciją.</p> <p>Jis teigė, kad laukia daugiau informacijos iš Škotijoje atlikto tyrimo autorių.</p> <p>„Mes niekada nekritikavome vakcinų“, - sakė jis, apibūdindamas ją kaip „labai gerą“.</p> <p>Vokietija nuo pandemijos pradžios patvirtino daugiau nei 2,4 milijono koronaviruso atvejų ir šiek tiek daugiau nei 70 000 mirčių.</p> <p>Užrakinimas galioja nuo gruodžio vidurio, tačiau infekcijos vis dar kelia nerimą ir pareigūnams buvo sunku sušvelninti apribojimus.</p> <p>Praėjusią savaitę paskelbti duomenys parodė, kad per pastarąsias septynias dienas infekcijos dažnis padidėjo iki 59,3 atvejo 100 000 žmonių, palyginti su 57 savaitė anksčiau.</p> <p>Vyriausybė turėtų susitikti kitą savaitę, kad apsvaistytų blokavimo taisykles, galiojančias iki kovo 7 d.</p> <p>Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų banga, jei per greitai bus panaikinta užraktas.</p>
--	--

Appendix 2. Experts' survey on acceptability

Link to the survey: <https://forms.gle/WvzugSVwW2zHtPzH6>

Priimtimumo tyrimas

Sveiki. Esu Greta Pinkevičiūtė, Kauno technologijos universiteto technikos kalbos vertimo ir lokalizacijos programos antro kurso magistro studentė. Atlieku tyrimą, kuriame vertinamas teksto priimtimumas. Šio klausimyno tikslas yra įvertinti mašininio vertimo suprantamumą, naudojamumą, patogumą.

Klausimynas anoniminis. Surinkti duomenys bus naudojami tik mokslinio tyrimo tikslais.

Perskaitę toliau pateikiamą straipsnio ištrauką ir pateikę atsakymus prisidėsite prie magistro baigiamojo projekto ruošimo.

Ačiū!

Kitas

1 dalis. Perskaitykite žemiau pateiktą straipsnio ištrauką ir įvertinkite žemiau pateiktus teiginius.

„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms. Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi. ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakciną naudoti visoms amžiaus grupėms. Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms. JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jį. „Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina“. Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.

Pagrindinė teksto mintis yra aiški. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Tekstas gramatiškai ir stilistiškai tvarkingas, suprantamas. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Šis tekstas yra naudingas, informatyvus. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Pasitikėčiau šaltiniu, kuriame šis tekstas yra pateiktas. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Tekstas skamba natūraliai. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Tekstas taisyklingas, jo kokybė puiki. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

2 dalis. Atsakykite į klausimus apie kiekvieną iš 7-ių pateiktų sakinių.

Atgal

Kitas

1 sakiny: „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti

Taip, lengvai

Taip, turbūt

Nei taip, nei ne

Sunku nustatyti

Nėra

Ar galite nustatyti, kur sakinyje yra klaida, jei ji yra.

Ar galite nustatyti, kur sakinyje yra klaida, jei ji yra?

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas

Ar taisytumėte klaidą sakinyje, jei ji yra?

Taip

Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

1 sakiny: „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.

Kritinė

Didelė

Nedidelė

Nereikšminga

Atgal

Kitas

2 sakiny: Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti ▼

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas _____

Ar taisytumėte klaidą sakinyje, jei ji yra?

Taip

Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

2 sakiny: Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.

Kritinė

Didelė

Nedidelė

Nereikšminga

Atgal

Kitas

3 sakiny: ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinos naudojimą visoms amžiaus grupėms.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti ▼

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas _____

Ar taisytumėte klaidą sakinyje, jei ji yra?

Taip

Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

3 sakiny: ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinos naudojimą visoms amžiaus grupėms.

- Kritinė
- Didelė
- Nedidelė
- Nereikšminga

Atgal

Kitas

4 sakiny: Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas

Ar taisytumėte klaidą sakinyje, jei ji yra?

- Taip
- Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

4 sakiny: Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.

- Kritinė
- Didelė
- Nedidelė
- Nereikšminga

Atgal

Kitas

5 sakiny: JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jeb.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti ▼

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas _____

Ar taisytumėte klaidą sakinyje, jei ji yra?

Taip

Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

5 sakins: JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jаб.

Kritisė

Didelė

Nedidelė

Nereikšminga

Atgal

Kitas

6 sakins: „Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.“

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti ▼

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas _____

Ar taisytumėte klaidą sakinyje, jei ji yra?

Taip

Ne

Jei sakinyje klaida yra, kokio ji sunkumo?

6 sakiny: „Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.“

- Kritinė
- Didelė
- Nedidelė
- Nereikšminga

Atgal

Kitas

7 sakiny: Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.

Ar galite nustatyti, kur sakinyje yra klaida? *

Pasirinkti

Nukopijuokite ir įklijuokite klaidą, jei ji yra.

Jūsų atsakymas

Ar taisytumėte klaidą sakinyje, jei ji yra?

- Taip
- Ne

Jeif sakinyje klaida yra, kokio ji sunkumo?

7 sakiny: Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.

- Kritinė
- Didelė
- Nedidelė
- Nereikšminga

Atgal

Kitas

Appendix 3. Non-experts' survey on acceptability

Link to the survey: <https://forms.gle/FLz3Ft4aC94zDcsTA>

Priimtino tyrimas

Sveiki. Esu Greta Pinkevičiūtė, Kauno technologijos universiteto technikos kalbos vertimo ir lokalizacijos programos antro kurso magistro studentė. Atlieku tyrimą, kuriame vertinamas teksto priimtumas. Šio klausimyno tikslas yra įvertinti mašininio vertimo suprantamumą, naudojamumą, patogumą.

Klausimynas anoniminis. Surinkti duomenys bus naudojami tik mokslinio tyrimo tikslais.

Perskaityt toliau pateikiamą straipsnio ištrauką ir pateikt atsakymus prisidėsite prie magistro baigiamojo projekto ruošimo.

Ačiū!

[Kitas](#)

Priimtino tyrimas

***Privaloma**

1 dalis. Perskaitykite žemiau pateiktą straipsnio ištrauką ir įvertinkite žemiau pateiktus teiginius.

„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms. Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi. ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakciną naudojimą visoms amžiaus grupėms. Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms. JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jeb. „Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina“. Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.

Pagrindinė teksto mintis yra aiški. *

1 – Visiškai sutinku

2 – Sutinku

3 – Nei sutinku, nei nesutinku

4 – Nesutinku

5 – Visiškai nesutinku

Tekstas gramatiškai ir stilistiškai tvarkingas, suprantamas. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Šis tekstas yra naudingas, informatyvus. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Pasitikėčiau šaltiniu, kuriame šis tekstas yra pateiktas. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Tekstas skamba natūraliai. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Tekstas taisyklingas, jo kokybė puiki. *

- 1 – Visiškai sutinku
- 2 – Sutinku
- 3 – Nei sutinku, nei nesutinku
- 4 – Nesutinku
- 5 – Visiškai nesutinku

Atgal

Kitas

1 sakiny: „Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijuokite ir įklijuokite neaiškų sakinyo segmentą (žodį, frazę).

Jūsų atsakymas _____

Atgal

Kitas

2 sakiny: Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijuokite ir įklijuokite neaiškų sakinyo segmentą (žodį, frazę).

Jūsų atsakymas _____

Atgal

Kitas

3 sakiny: ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinos naudojimą visoms amžiaus grupėms.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijukite ir įklijuokite neaiškų sakinio segmentą (žodį, frazę).

Jūsų atsakymas _____

Atgal

Kitas

4 sakiny: Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijukite ir įklijuokite neaiškų sakinio segmentą (žodį, frazę).

Jūsų atsakymas _____

Atgal

Kitas

5 sakiny: JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jab.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijuokite ir įklijuokite neaiškų sakinyo segmentą (žodį, frazę).

Jūsų atsakymas

Atgal

Kitas

6 sakiny: „Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina.“

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijuokite ir įklijuokite neaiškų sakinyo segmentą (žodį, frazę).

Jūsų atsakymas

Atgal

Kitas

7 sakiny: Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.

Ar sakiny suprantamas? *

- Taip
- Ne
- Dalinai

Jeį pažymėjote "Ne" arba "Dalinai", nukopijuokite ir įklijuokite neaiškų sakinio segmentą (žodį, frazę).

Jūsų atsakymas

Atgal

Kitas

Appendix 4. Machine translation quality assessment by employing Multidimensional Quality Metrics

Errors of the Accuracy dimension					
	Source sentence	Target sentence	2nd level subtype	3rd level subtype (if applicable)	Error weight
	Covid vaccine: Germany urged to back AstraZeneca jab for over-65s	„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms	Mistranslation	Overly literal	1
	Covid vaccine: Germany urged to back AstraZeneca jab for over-65s	„Covid“ vakcina: Vokietija paragino paremti „AstraZeneca“ žandikaulį vyresniems nei 65 metų žmonėms	Mistranslation		100
	A senior German immunologist has urged his country to change its mind and start allowing over-65s to receive the Oxford-AstraZeneca vaccine .	Vyresnysis vokiečių imunologas paragino savo šalį persigalvoti ir leisti leisti vyresniems nei 65 metų asmenims skiepytis „Oxford-AstraZeneca“.	Omission		1
	In a BBC interview , Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.	BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą vyresnio amžiaus žmonių.	Under translation		0
	In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.	BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą vyresnio amžiaus žmonių.	Mistranslation	Overly literal	1
	In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.	BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių.	Omission		10
	Carsten Watzl urged Angela Merkel to have the vaccine live on TV to prove it is safe.	Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.	Mistranslation	Overly literal	10
	Prof Watzl's call comes after recent studies in Scotland showed the AstraZeneca jab to be effective among the elderly.	Prof. Watzlio kvietimas įvyko po to, kai neseniai Škotijoje atlikti tyrimai parodė, kad „AstraZeneca“ žandikaulis yra veiksmingas vyresnio amžiaus žmonėms.	Mistranslation	Overly literal	0

Prof Watzl's call comes after recent studies in Scotland showed the AstraZeneca jab to be effective among the elderly.	Prof. Watzlio kvietimas įvyko po to, kai neseniai Škotijoje atlikti tyrimai parodė, kad „AstraZeneca“ žandikaulis yra veiksmingas vyresnio amžiaus žmonėms.	Mistranslation		100
The country is currently struggling to avoid a third wave of infections as cases remain stubbornly high.	Šiuo metu šalis stengiasi išvengti trečios infekcijų bangos, nes atvejų vis dar atkakliai daug.	Mistranslation	Overly literal	1
But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Omission		10
But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Mistranslation		10
But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Mistranslation	Overly literal	1
But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Mistranslation		100
They cited insufficient data on its efficacy for older people.	Jie nurodė nepakankamus duomenis apie jo veiksmingumą vyresnio amžiaus žmonėms.	Mistranslation	Ambiguous translation	0
German health authorities have so far used fewer than 300,000 of the 1.17 million doses of the AstraZeneca vaccine the country has received.	Vokietijos sveikatos priežiūros institucijos kol kas sunaudojo mažiau nei 300 000 iš šalyje gautų 1,17 milijonų vakcinos „AstraZeneca“ dozių.	Mistranslation	Overly literal	0
In January, French President Emmanuel Macron said the jab was "quasi-ineffective" for older age groups - a claim strongly refuted by the UK government and British medical regulators.	Sausio mėn. Prancūzijos prezidentas Emmanuelis Macronas teigė, kad vyresnio amžiaus žmonėms šiukšlės buvo „beveik neveiksmingos“ - JK vyriausybė ir Didžiosios Britanijos medicinos	Mistranslation		100

		priežiūros institucijos tvirtai paneigė šį teiginį.			
	AstraZeneca itself says the vaccine is effective at all ages.	Pati „AstraZeneca“ teigia, kad vakcina yra veiksminga bet kokio amžiaus.	Mistranslation	Overly literal	0
	AstraZeneca itself says the vaccine is effective at all age groups .	Pati „AstraZeneca“ teigia, kad vakcina yra veiksminga bet kokio amžiaus .	Under translation		1
	The UK is among countries that have approved the jab for all age groups.	JK yra tarp šalių, kurios patvirtino visų amžiaus grupių jab .	Untranslated		100
	The decision was boosted by recent research led by Public Health Scotland , which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.	Mistranslation	Should not have been translated	1
	The decision was boosted by recent research led by Public Health Scotland, which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.	Mistranslation	Overly literal	1
	The decision was boosted by recent research led by Public Health Scotland, which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.	Untranslated		0
	The decision was boosted by recent research led by Public Health Scotland, which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių .	Mistranslation		100
	Among the over 80s, there was an overall 81% reduction in the numbers admitted to hospital when the results for both vaccines were combined.	Tarp vyresnių nei 80 metų pacientų, paguldytų į ligoninę, skaičius sumažėjo 81% , kai buvo susieti abiejų vakcinų rezultatai.	Untranslated		0

Among the over 80s, there was an overall 81% reduction in the numbers admitted to hospital when the results for both vaccines were combined.	Tarp vyresnių nei 80 metų pacientų, paguldytų į ligoninę, skaičius sumažėjo 81%, kai buvo susieti abiejų vakcinų rezultatai.	Mistranslation	Overly literal	1
"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą"- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Mistranslation	Should not have been translated	1
"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą "- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Mistranslation	Overly literal	0
"In order for us to reach our vaccination goals we need people to get this vaccine.	„Kad galėtume pasiekti užsibrėžtus skiepavimo tikslus, reikia, kad žmonės gautų šią vakciną.	Mistranslation	Overly literal	0
"We do know that the vaccine works in that age group.	"- Mes tikrai žinome, kad vakcina veikia toje amžiaus grupėje.	Mistranslation	Overly literal	1
Prof Watzl predicted German regulators would "reverse course based on the data coming out of Scotland".	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.	Mistranslation		1
Prof Watzl predicted German regulators would "reverse course based on the data coming out of Scotland".	Profas Watzlas prognozavo, kad Vokietijos reguliotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.	Mistranslation	Overly literal	1
Prof Watzl predicted German regulators would " reverse course based on the data coming out of Scotland".	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „ pakeis kursą , remdamiesi iš Škotijos gaunamais duomenimis“.	Mistranslation	Overly literal	0
" If at that point [Chancellor] Angela Merkel were to go on live television and have the vaccine, that would be great," he added.	„ Jei tuo metu [kanclerė] Angela Merkel eitų į tiesioginę televiziją ir turėtų vakciną, būtų puiku“, - pridūrė jis.	Mistranslation	Ambiguous translation	0
"If at that point [Chancellor] Angela Merkel were to go on live television and have the vaccine , that would be great," he added.	„Jei tuo metu [kanclerė] Angela Merkel eitų į tiesioginę televiziją ir turėtų vakciną , būtų puiku“, - pridūrė jis.	Mistranslation	Overly literal	10
On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated recommendation" on the AstraZeneca vaccine.	Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą,, AstraZeneca “vakcinos rekomendaciją.	Mistranslation		1

"We have never criticised the vaccine," he said, describing it as "very good" .	„Mes niekada nekritikavome vakcinų“, - sakė jis, apibūdinamas ją kaip „labai gerą	Mistranslation	Overly literal	0
A lockdown has been in place since mid-December but infections are still worryingly high and officials have found it difficult to ease restrictions.	Užrakinimas galioja nuo gruodžio vidurio, tačiau infekcijos vis dar kelia nerimą ir pareigūnams buvo sunku sušvelninti apribojimus.	Mistranslation	Overly literal	10
A lockdown has been in place since mid-December but infections are still worryingly high and officials have found it difficult to ease restrictions.	Užrakinimas galioja nuo gruodžio vidurio, tačiau infekcijos vis dar kelia nerimą ir pareigūnams buvo sunku sušvelninti apribojimus.	Mistranslation	Overly literal	1
A lockdown has been in place since mid-December but infections are still worryingly high and officials have found it difficult to ease restrictions.	Užrakinimas galioja nuo gruodžio vidurio, tačiau infekcijos vis dar kelia nerimą ir pareigūnams buvo sunku sušvelninti apribojimus.	Mistranslation	Overly literal	1
Figures released last week showed the infection rate had risen to 59.3 cases per 100,000 people over the past seven days, from 57 a week earlier.	Praėjusią savaitę paskelbti duomenys parodė, kad per pastarąsias septynias dienas infekcijos dažnis padidėjo iki 59,3 atvejo 100 000 žmonių, palyginti su 57 savaitė anksčiau.	Mistranslation	Overly literal	1
Figures released last week showed the infection rate had risen to 59.3 cases per 100,000 people over the past seven days, from 57 a week earlier.	Praėjusią savaitę paskelbti duomenys parodė, kad per pastarąsias septynias dienas infekcijos dažnis padidėjo iki 59,3 atvejo 100 000 žmonių, palyginti su 57 savaitė anksčiau.	Mistranslation	Overly literal	1
The government is due to meet next week to consider lockdown rules that run to 7 March.	Vyriausybė turėtų susitikti kitą savaitę, kad apsvarstytų blokavimo taisykles , galiojančias iki kovo 7 d.	Mistranslation	Overly literal	10
Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.	Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų bangą , jei per greitai bus panaikinta užraktas.	Mistranslation	Overly literal	1
Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.	Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų banga, jei per greitai bus panaikinta užraktas .	Mistranslation	Overly literal	10
			Total:	699

Errors of the Fluency dimension

Source sentence	Target sentence	2 nd level subtype	3 rd level subtype (if applicable)	Error weight
-----------------	-----------------	-------------------------------	---	--------------

	A senior German immunologist has urged his country to change its mind and start allowing over-65s to receive the Oxford-AstraZeneca vaccine.	Vyresnysis vokiečių imunologas paragino savo šalį persigalvoti ir leisti leisti vyresniems nei 65 metų asmenims skiepytis „Oxford-AstraZeneca“.	Duplication		0
	In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people .	BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių .	Grammar	Word form (caused by the Mistranslation issue of the Accuracy dimension (Appendix X))	1
	Germany's vaccine commission is currently reviewing its recommendation.	Vokietijos vakcinų komisija šiuo metu peržiūri jos rekomendaciją.	Ambiguity	Unclear reference	1
	Germany is one of several EU states that have expressed doubts over the efficacy of the vaccine in older people.	Vokietija yra viena iš kelių ES valstybių, kurios išreiškė abejones dėl vakcinės veiksmingumo vyresnio amžiaus žmonėms.	Grammar	Word form	0
	The EU's medical regulator approved the use of the Oxford-AstraZeneca vaccine for all age groups in January.	ES medicinos priežiūros institucija sausio mėn. Patvirtino „Oxford-AstraZeneca“ vakcinės naudojimą visoms amžiaus grupėms.	Spelling	Capitalisation	1
	But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Grammar	Word form	10
	But the rollout was met by some public scepticism after regulators in countries including France, Germany and Italy recommended that it should not be used for people over 65.	Tačiau viešai skeptiškai reagavo į skleidimą, kai reguliuotojai šalyse, įskaitant Prancūziją, Vokietiją ir Italiją, rekomendavo jo naudoti vyresniems nei 65 metų žmonėms.	Ambiguity	Unclear reference	1
	They cited insufficient data on its efficacy for older people.	Jie nurodė nepakankamus duomenis apie jo veiksmingumą vyresnio amžiaus žmonėms.	Ambiguity	Unclear reference	1
	German health authorities have so far used fewer than 300,000 of the 1.17 million doses of the AstraZeneca vaccine the country has received.	Vokietijos sveikatos priežiūros institucijos kol kas sunaudojo mažiau nei 300 000 iš šalyje gautų 1,17 milijonų vakcinės „AstraZeneca“ dozių .	Grammar	Word order	0
	The decision was boosted by recent research led by Public Health Scotland, which found that four weeks	Šį sprendimą paskatino naujaisi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus	Coherence		0

	after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.			
	The decision was boosted by recent research led by Public Health Scotland, which found that four weeks after the first dose, hospital admissions were reduced by 85% and 94% for the Pfizer-BioNTech and AstraZeneca jabs respectively.	Šį sprendimą paskatino naujausi Škotijos visuomenės sveikatos vadovaujami tyrimai, kurie parodė, kad praėjus keturioms savaitėms po pirmosios dozės ligoninių priėmimas sumažėjo atitinkamai 85% ir 94% „Pfizer-BioNTech“ ir „AstraZeneca“ žandikaulių.	Grammar	Word order	1
	The recent data from Scotland clearly show it elicits an immune response, the elderly are protected from severe disease by this vaccine."	Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina."	Ambiguity	Unclear reference	1
	The recent data from Scotland clearly show it elicits an immune response, the elderly are protected from severe disease by this vaccine. "	Naujausi Škotijos duomenys aiškiai rodo, kad tai sukelia imuninį atsaką, pagyvenusius žmones apsaugo nuo sunkių ligų ši vakcina. "	Grammar	Word order	0
	On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated recommendation" on the AstraZeneca vaccine.	Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą,, AstraZeneca “vakcinos rekomendaciją.	Typography		1
	On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated recommendation" on the AstraZeneca vaccine.	Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą,, AstraZeneca “vakcinos rekomendaciją.	Typography	Unpaired quote marks or brackets	1
	A lockdown has been in place since mid-December but infections are still worryingly high and officials have found it difficult to ease restrictions.	Užrakinimas galioja nuo gruodžio vidurio, tačiau infekcijos vis dar kelia nerimą ir pareigūnams buvo sunku sušvelninti apribojimus	Grammar	Word form (tense)	1
	Figures released last week showed the infection rate had risen to 59.3 cases per 100,000 people over the past seven days, from 57 a week earlier.	Praėjusių savaitę paskelbti duomenys parodė, kad per pastarąsias septynias dienas infekcijos dažnis padidėjo iki 59,3 atvejo 100 000 žmonių, palyginti su 57 savaitę anksčiau.	Grammar	Word form	0

Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.	Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų banga, jei per greitai bus panaikinta užraktas.	Grammar	Word form	1
			Total:	21

Errors of the Style dimension

Source sentence	Target sentence	2 nd level subtype	3 rd level subtype (if applicable)	Error weight
"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą "- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Unidiomatic		0
Prof Watzl predicted German regulators would reverse course based on the data coming out of Scotland".	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „ pakeis kursą “, remdamiesi iš Škotijos gaunamais duomenimis“.	Unidiomatic		0
Chancellor Merkel has warned that the country could be hit by a third wave of cases if the lockdown is lifted too quickly.	Kanclerė Merkel perspėjo, kad šalį gali ištikti trečioji atvejų banga, jei per greitai bus panaikinta užraktas.	Unidiomatic		0
			Total:	0

Errors of the Locale convention dimension

Source sentence	Target sentence	2 nd level subtype	3 rd level subtype (if applicable)	Error weight
In a BBC interview, Carsten Watzl, head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.	BBC interviu Carstenas Watzlas, Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių.	Locale-specific punctuation		0
In a BBC interview, Carsten Watzl , head of the German Society for Immunology, predicted regulators would have to reverse their decision to not recommend the jab for older people.	BBC interviu Carstenas Watzlas , Vokietijos imunologijos draugijos vadovas, prognozavo, kad reguliavimo institucijos turės pakeisti savo sprendimą nerekomenduoti vyresnio amžiaus žmonių.	Name format		1
Carsten Watzl urged Angela Merkel to have the vaccine live on TV to prove it is safe.	Carstenas Watzlas paragino Angelą Merkel vakciną tiesiogiai transliuoti per televiziją, kad įrodytų, jog ji yra saugi.	Name format		1
Prof Watzl's call comes after recent studies in Scotland showed the	Prof. Watzlio kvietimas įvyko po to, kai neseniai Škotijoje atlikti tyrimai	Name format		1

	AstraZeneca jab to be effective among the elderly.	parodė, kad „AstraZeneca“ žandikaulis yra veiksmingas vyresnio amžiaus žmonėms.			
	In January, French President Emmanuel Macron said the jab was "quasi-ineffective" for older age groups - a claim strongly refuted by the UK government and British medical regulators.	Sausio mėn. Prancūzijos prezidentas Emmanuelis Macronas teigė, kad vyresnio amžiaus žmonėms šiukšlės buvo „beveik neveiksmingos“ - JK vyriausybė ir Didžiosios Britanijos medicinos priežiūros institucijos tvirtai paneigė šį teiginį.	Name format		1
	In January, French President Emmanuel Macron said the jab was "quasi-ineffective" for older age groups - a claim strongly refuted by the UK government and British medical regulators. AstraZeneca itself says the vaccine is effective at all ages.	Sausio mėn. Prancūzijos prezidentas Emmanuelis Macronas teigė, kad vyresnio amžiaus žmonėms šiukšlės buvo „beveik neveiksmingos“ - JK vyriausybė ir Didžiosios Britanijos medicinos priežiūros institucijos tvirtai paneigė šį teiginį.	Locale-specific punctuation		0
	"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą"- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Locale-specific punctuation		0
	"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą" - sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Name format		1
	"I think Germany will also reverse course soon," Prof Watzl told BBC Radio 4's Broadcasting House programme.	"- Manau, kad Vokietija taip pat greitai pakeis kursą"- sakė prof. Watzlas BBC radijo 4 laidos "Broadcasting House" programai.	Locale-specific punctuation		0
	"We do know that the vaccine works in that age group.	"- Mes tikrai žinome, kad vakcina veikia toje amžiaus grupėje.	Locale-specific punctuation		0
	"If at that point [Chancellor] Angela Merkel were to go on live television and have the vaccine, that would be great," he added.	„Jei tuo metu [kanclerė] Angela Merkel eitų į tiesioginę televiziją ir turėtų vakciną, būtų puiku“, - pridūrė jis.	Locale-specific punctuation		0
	Prof Watzl predicted German regulators would "reverse course based on the data coming out of Scotland".	Profas Watzlas prognozavo, kad Vokietijos reguliuotojai „pakeis kursą“, remdamiesi iš Škotijos gaunamais duomenimis“.	Name format		1
	On Friday, Thomas Mertens, who heads Germany's vaccine commission, told broadcaster ZDF that the body would "very soon publish a new updated	Penktadienį Thomasas Mertensas, vadovaujantis Vokietijos vakcinų komisijai, transliuotojui ZDF sakė, kad organizmas „labai greitai paskelbs naują atnaujintą,,	Locale-specific punctuation		0

	recommendation" on the AstraZeneca vaccine.	AstraZeneca "vakcinu rekomendaciją.			
	"We have never criticised the vaccine," he said, describing it as "very good".	„Mes niekada nekritikavome vakcinų“, sakė jis, apibūdindamas ją kaip „labai gerą“	Locale-specific punctuation		0
				Total:	6