Human Evaluation of Machine-Translated Texts

Čanžar, Margareta

UNIVERSITY OF RIJEKA

FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Margareta Čanžar

Human Evaluation of Machine-Translated Texts

Master's thesis

Rijeka, 2024

UNIVERSITY OF RIJEKA

FACULTY OF HUMANITIES AND SOCIAL SCIENCES

DEPARTMENT OF ENGLISH LANGUAGE AND LITERATURE

DIVISION OF TRANSLATION STUDIES

Margareta Čanžar

0009080741

Human Evaluation of Machine-Translated Texts

Master's thesis

Graduate Study Programme in Translation Studies

Supervisor: Mirjana Borucinsky, PhD

Rijeka, May 2024

Izjava o autorstvu

Ovime potvrđujem da sam osobno napisala diplomski rad pod naslovom *Human Evaluation of Machine-Translated Texts* i da sam njegova autorica.

Svi dijelovi rada, podaci ili ideje koje su u radu citirane ili se temelje na drugim izvorima (mrežnim izvorima, literaturi i drugom), u radu su jasno označeni kao takvi te su navedeni u popisu literature.

Margareta Čanžar

Table of Contents

Abstract

The rapid progressions in machine translation technologies have sparked significant interest in comparing their outputs against those delivered by human translators. Therefore, this Master's thesis presents quantitative and qualitative research on the topic of evaluation of translation, specifically, the evaluation of human and machine translation for the Croatian-English language pair in domain-specific terminology. Even though machine translation has significantly improved in the last decade, there is still need to evaluate its quality. Therefore, this thesis explores the quality of machine-translated texts, i.e. if such texts are considered acceptable by native English-speaking evaluators, and if non-native English speakers were able to identify which phrase is a machine translation. The research was done through a survey with multiple choice questions; one survey was made for native speakers of English, and one for non-native speakers of English (i.e. native Croatian speakers), but the same phrases were given in both surveys. The results of the survey conducted with native speakers reveal that native English speakers find human translations more acceptable than the translations done by a machine. The results of the survey conducted with non-native speakers reveal that native Croatian speakers are able to distinguish between machine-translated and human-translated phrases in English.

Keywords: *evaluation, human translation, machine translation, survey, English, Croatian*

1. Introduction

In the contemporary age, translation is marked by the coexistence of human translators and machine translation systems that are improving daily due to the rapid development of artificial intelligence.

As technology has been rapidly advancing, machine translation has become a means to bridge the gap between languages. This allows quickly accessible translations for different purposes, ranging from personal to professional documents. However, the question that is posed when it comes to human versus machine translation evaluation is how they compare to each other in terms of overall quality, accuracy and fluency. Even though machines are able to produce and process a large number of data in the smallest amount of time possible, they typically cannot compete with the ability of humans to capture meaning, idiomatic expressions and specific terminology, as well as culturally marked concepts.

The importance of evaluating machine translation systems lies in assessing their usability and effectiveness in real-world context.

Therefore, this Master's thesis presents the results of the evaluation of machine versus human translation, specifically in rendering Croatian source language phrases into English. Croatian is underrepresented and considered a "small" language (as opposed to English), therefore making it important to research how well the machine translation systems are able to decode and convey the meaning of Croatian phrases into another language. Specifically in this survey, into English, a language that is still the most represented in the translation databases and translation memories, considering its usage and influence all over the world.

The study presented in this Master's thesis aims to evaluate machine-translated phrases and to compare them against human-translated phrases. This evaluation is not based on metrics but is rather a qualitative evaluation performed by native and non-native speakers of English. One of the main aims was to find out if native speakers of English will recognize machine-translated phrases as such.

The first part of the thesis presents a theoretical background, i.e. it provides some general information on machine translation, its benefits and disadvantages, and the difficulties of using

machine translation systems for translating languages such as Croatian. Certain methods for evaluating translation and the differences between machine and human evaluation are presented as well. Then, the methodology for this survey is presented, i.e. the aim of the research, research questions, research method, description of evaluators, results, followed by discussion. Finally, at the end of the thesis, the conclusion summarizes the main points and provides some suggestions for future research.

2. Theoretical Background

2.1.Machine Translation

Human translation is costly and time-consuming when it comes to meeting the demands of the translation industry and the use of machine translation presents an answer to dealing with such an issue. Machine translation (MT) is automated translation that should not be confused with the term computer-assisted (computer-aided) translation (CAT), because they are not the same. The former refers to the process during which a computer software is used in order to translate a text from the source language into the target language (Okpor, 2014). Machine translation does not include any human intervention but is essentially just "uploading" the text that one wishes to translate into a software. Machine translation takes advantage of the computer's capacity to calculate in order to assess a statement or a sentence's structure in the source language, and then break it down into pieces that are easily translatable. Afterwards, it generates a statement with the same structure in the target language. Machine translation uses extensive multilingual dictionaries and pre-translated corpora of text (Craciunescu et al., 2007).

On the other hand, computer-assisted translation is a term that refers to a tool that translators use to help them translate a text into a target language (Han, 2020). Due to CAT responding more realistically to the needs of translating, it has become more popular than machine translation. CAT helps translators to work more swiftly and accurately with a number of tools that it offers. The most important of those tools are translation memories and terminology databases. One of the biggest advantages of using CAT tools is working with a digital document, which means that translators are provided with nonconsecutive access to information that can be used according to their needs. It becomes simple to do tasks like word and sentence analysis, and context verification of text usage in phrases inside the source text (Craciunescu et al., 2007). However, such translation systems are still not capable of creating a text that can be immediately used. Many CAT tools nowadays include machine translation, so the distinction between the two is becoming less transparent.

Because it facilitates communication and transfer of meaning across languages, machine translation is gaining popularity with an increased speed as a study area in the fields of computer

science, computational linguistics, and information and communication sciences. It can also be quite limited and should not be considered a means to replace humans, but to speed up the translation process and save time. It is important to note that Croatian is usually considered to be low-resourced, specifically when it comes to technology and service available, due to it being underrepresented (Dunđer, 2020). Though machine translation's rapid advancement has greatly improved translation quality, not all language pairs respond well to machine translation techniques. Morphologically rich languages pose challenges for machine translation, particularly when translating from a morphologically simple to a morphologically complex language. It is important to create morphological differences in the target language that are absent in the source language. The development of morphology-aware approaches frequently depends on language-specific tools, but these are not always accessible (Sepesy Maučec and Donaj, 2019). A lot of morphologically rich languages fall in the category of low-resource languages. A group of morphologically rich languages is a group of highly inflected languages. These are challenging for language technology applications as well as for machine translation. The primary issue in highly inflected languages is that a large number of inflected word forms causes data sparsity, which leads to inaccurate estimates in statistical machine translation. The majority of words in a given corpus only appear a few times at most. Some approaches use modeling units other than words, such as stems and ends, lemmas and morphosyntactic tags, etc., in an attempt to lessen the issue of data sparsity. Another issue in inflectional languages is the free word order. The source sentence typically provides very little information about the target word order. To make the words on the source side appear closer to their final places on the target side, pre-ordering approaches learn to preprocess the source phrase during training. Pronouns pose another issue for inflectional languages and are thus frequently translated incorrectly. Additionally, there are numerous instances in inflectional languages where the subject is omitted entirely. Differences in how negation is expressed might also be problematic. Slavic languages are extremely inflected languages that pose a great deal of challenges for machine translation (Sepesy Maučec and Donaj, 2019).

However, machine translation systems are currently available for widely spoken languages, but they are not as advanced for less developed languages (e.g. Croatian) (Dunđer, 2020). English dominates the “translation market”, due to machine translation developing based on supply and demand (Craciunescu et al., 2007). In order to achieve and reach the goals of the translation market, machine translation combined with human post-editing is being used more frequently (O’Brien et

al., 2014; Ahrenberg, 2017). This also means that the expectations for machine translation have been set higher, as well as the quality of the translation it produces.

2.2.Human Translation and Evaluation of MT Systems

Unlike machine translation, human translation frequently aims at higher quality, creating texts that meet target cultural language conventions and are tailored to the readers' presumptions (Ahrenberg, 2017). It is for that reason that human translation is taken as the golden standard against which machine translation is evaluated.

Escribe (2019) states that the issues with human evaluation are that it can be very time-consuming, quite expensive and, normally, not reusable. Also, these studies are highly subjective. This is due to the fact that human evaluators do not necessarily agree on the quality of the machine translation output. According to Lee et al. (2023), some other challenges that human evaluation faces are inter-annotator agreement, and the fact that human evaluators can bring presumptions and their own biases to the process of evaluation. This may lead to some inconsistent results. Moreover, a very challenging task is error categorization. Error analysis and categorization serve as a basis for identifying what kind of errors were produced by the system and if they can be eliminated. Error analysis is typically used in the translation industry for evaluation, as it can provide several solutions to improve the system, such as enhanced comprehension of human or artificial agent performance. However, it takes a lot of time and demands in-depth expertise of the annotators (Popović, 2018; Munkova et al., 2021). The exact process of error categorization and analysis depends on the language. Usually, human evaluators will search for errors such as “missing words,” “incorrect word order,” “added words,” or “wrong part of speech.” (“Machine Translation 101 – Part 3,” 2021).

Nonetheless, human judgment is crucial for creating effective evaluation systems and deciphering the results they provide. Since human analyses frequently serve as a framework for machine-translation tools, human input is essential to improve machine translation evaluation systems. Although automatic metrics such as BLEU are commonly used to analyze machine translation systems, human assessors' judgment of machine translation output is considered the gold standard for assessing quality. Evaluation should detect the difference in quality between

machine translation and gold standard human translation (meaningfulness) (Licht et al., 2022). The concept of translation quality is subjective in general. The most widely recognized measure, or "Gold Standard", involves having human judges who are bilingual to assess the translation by comparing the inputs in the source language with the outputs in the target language and then rating the translation according to their subjective assessment. The most common format for these evaluations is a multi-point scale (Sanders et al., 2010).

3. Evaluation of Machine Translation

Research on the evaluation of machine translation systems is essential to maximize the performance of such systems, but also to assess how successful current machine translation systems are. According to Dorr et al. (2011), there are two different approaches for evaluating machine translation: *Glass Box* evaluation and *Black Box* evaluation. *Glass Box* evaluation refers to measuring the system's quality using the internal properties of the system and the primary focus is on the linguistic coverage of the system and the theories applied to the linguistic phenomena. However, *Black Box* evaluation disregards the internal mechanisms and measures the system's quality by concentrating only on the output. Within the black box evaluation, we differ *intrinsic* and *extrinsic* measures to examine the accuracy and usefulness of the machine translation product.

3.1.Intrinsic Measures

Intrinsic measures evaluate the quality of machine translation output. They do this by making a comparison with a set of good-quality reference translations. We also differentiate human intrinsic measures and automatic intrinsic measures. Human intrinsic measures rely on subjective assessments of fluency and adequacy. Automatic intrinsic measures, however, use a simple sentence similarity measure to rate machine translation systems against reference translations. Some of the automatic metrics are: BLEU (Bilingual Evaluation Understudy), NIST (National Institute of Standards and Technology), METEOR (Metric for Evaluation of Translation with Explicit Ordering), WER (Word Error Rate), PER (Position-independent Error Rate), GTM (General Text Matcher), TER (Translation Error Rate), and CDER (Cover Disjoint Error Rate). The use of these metrics has advanced machine translation optimization, specifically in programs

like GALE. Automatic metrics are essential in machine translation research because they allow rapid testing of novel features and models, and they make automatic optimization of system parameters possible (Dorr et al., 2011).

3.2.Human Judgment-based Evaluation

The machine translation community regularly uses a variety of human judgment-based evaluation techniques. The quality of system output can be evaluated directly in some situations, like when human judgments are used; indirectly in other situations, like when reading assessments or such tasks are conducted using the system output; and finally, like when the amount of work needed to rectify the system output is computed. Fluency and adequacy judgments are the most common human evaluation metrics (Callison-Burch, 2007; Dorr et al., 2011).

Fluency needs a speaker that is fluent in the target language to evaluate if the system output is fluent, regardless of whether the content of the output is an exact translation of the source words. Adequacy evaluates whether the important information in the source can be recovered from the system output, ignoring the fluency of the system output to the greatest extent possible. The annotator has to be bilingual in both (source and target) languages to be able to judge if the information is preserved in the translation. This makes the requirements for an annotator of adequacy stricter than the ones for fluency. Usually, an annotator who is only fluent in the target language may also annotate adequacy by using a set of high-quality human translations of the original text. Each sentence in the system output is assessed independently for both adequacy and fluency, and evaluations are often given on a five or seven point scale (Przybocki, 2008; Dorr et al., 2011). Sometimes, these are averaged to give a single numerical score to a system output. There are studies that have shown poor correlation between annotators using this method, which poses a question of how reliable this method actually is (Snover, 2006; Dorr et al., 2011).

Nonetheless, human evaluation has served as a standard against which evaluation metrics are often judged. Judgments of semantic adequacy, as well as concepts such as understandability or fluency (Nubel, 1997; Dorr et al., 2011) have continued to be used as valuable benchmarks for the performance of machine translation systems and proposed machine translation metrics, despite the challenges in ensuring the reliability of human judgments (Turian, 2003; Dorr et al., 2011).

Using professional human translators for evaluation gives the best results when it comes to measuring quality and analyzing errors. It makes it easier to evaluate the key metrics (e.g. adequacy and fluency scores, post-editing measures, human ranking of translations at the sentence level, and task-based evaluations), although this appears to be expensive and time-consuming ("Machine Translation 101 – Part 3," 2021).

There are several ways by which human translators evaluate machine translation. One of them is to evaluate by giving a rating to the quality of the target translation. In such instance, a scale of 1-10 (or a percentage) is usually used, that ranges from "very bad quality" to "flawless quality." ("Machine Translation 101 – Part 3," 2021).

Another method for evaluation is assessing the adequacy of the machine translation. This refers to how much the meaning of the source text has been retained in the target text. It is typically rated on a scale from "no meaning retained" to "all meaning retained." Evaluation by using adequacy demands the fluency of human evaluators in both source and target language.

Another useful metric for human translators is fluency which is used to judge the quality of a translation. Scales for fluency typically go from "incomprehensible" to "fluent." This type of evaluation involves just the target text, meaning that the human translator does not need to know both languages. Due to the subjective nature of human judgment, it may be difficult to achieve a good level of intra-rater and inter-rater agreement (i.e., consistency of the same human evaluator and the consistency of multiple human evaluators). Therefore, standardized metrics for human evaluation do not exist as such ("Machine Translation 101 – Part 3," 2021).

However, in reality, various human evaluators use quite varied criteria to evaluate machine translation output, based on factors such as their language proficiency, experience to machine translation output, expectations regarding translation quality, and how the source text is presented. This is particularly problematic when the objective is to get meaningful ratings between language pairings, such as when determining if a machine translation system is good enough to be used for a particular language pair (Licht et al., 2022).

3.3.Automatic Evaluation Metrics

Unlike human evaluation, automatic metrics seem to be the most efficient solution due to their objectiveness and being quick and inexpensive. Automatic evaluation metrics work by comparing overlapping words. However, the issue with this is that metrics only take into account lexical similarity, but they do not actually measure sentence structure and grammatical, semantic diversity. An example of this would be that BLEU usually does not do good with sentences that appear to be similar semantically, but they have different structure and vocabulary (Lee et al., 2023).

As mentioned previously, the most popular automatic metric is BLEU. According to Lin and Och (2004, p.1), "the main idea of BLEU is to measure the translation closeness between a candidate translation and a set of reference translations with a numerical metric." BLEU is able to give very accurate results, which are correlated with human judgment (Papineni et al., 2002; Escribe, 2019).

3.4.Extrinsic Measures

Extrinsic measures, also known as task-based measures, are used to test how well the machine translation output performs a specific task, i.e. they measure how an automated process or system performs on degraded MT output (Popescu-Belis, 2007; Babych and Hartley, 2008).

Extrinsic, or task-based measures, are able to work with no human reference translation and it was originally suggested for human evaluation (Hutchins and Somers, 1992; Babych and Hartley, 2008). Task-based evaluation methods look at the structure and interaction of specific levels by taking an external view, i.e. they use functional models for machine translation quality (unlike BLUE, e.g., that uses structural models). Extrinsic measures evaluate how the text or specific contexts carry out a function that is outside of the structure. This means that such evaluation might be able to detect structural degradation at any level which contributes to this function. Task-based measures do not make explicit assumptions about certain combinations of structural features that carry out external textual functions (Babych and Hartley, 2008).

Dorr et al. (2011) give two examples of this, which are extrinsic measures of human performance based on document exploitation task accuracy, and measurements of human reading comprehension of machine-translated texts. These measures focus on testing the usefulness of machine translation.

As machine translation systems continue to advance, it is becoming more and more important to employ assessment techniques to examine the translations produced by such systems in order to create more effective systems. This is why the evaluation of machine translation developed into a field of its own (Escribe, 2019).

4. Methodology

This chapter presents the methodology of this research study, i.e. how the study was conducted. It explains the aim of the research and the research questions.

4.1.Aim

The aim of this research was to evaluate machine-translated phrases and to compare them against human-translated phrases. This evaluation is not based on metrics but is rather a qualitative evaluation performed by native and non-native speakers of English. It is believed that native speakers can judge the appropriateness of certain phrases in their L1, and some non-native speakers that participated in the study either possess a degree in English language and literature and/or Translation studies, or have some experience in translating.

The aims were to find out if native speakers of English regard machine-translated phrases as acceptable as the human-translated ones, and whether non-native speakers of English are able to distinguish between machine-translated and human-translated phrases in English.

4.2.Research Questions

1. Are native speakers of English able to distinguish between machine-translated and human-translated phrases in English? Do they perceive machine-translated phrases as acceptable as the human-translated phrases?
2. Are non-native speakers of English able to distinguish between machine-translated and human-translated phrases in English?
3. Given the improvement of machine translation systems, is the boundary between natural (human) and machine-translated language becoming less clear? OR is it becoming increasingly difficult to distinguish machine-translated phrases (texts) from human produced translations?

4.3.Research Method

The first step of this research was to compile parallel corpora. The corpora include:

1. A corpus of four original texts in Croatian;
2. A corpus of the four texts translated from Croatian into English by a human translator;
3. A corpus of the four texts translated from Croatian into English by machine.

The original texts in Croatian (i.e. source texts) were obtained from the Portal of Croatian scientific and professional journals, specifically from the Croatian Journal of Education's 2022 issue. The texts deal with topics regarding the fields of Education and Psychology, including research on students' attitude toward reading and reading assignments, gender (in)equality in child-rearing and housework between mothers and fathers from the perspective of children, gifted students with disabilities, and comparison of musically educated, athletically active, and other adolescents. All texts are written in a formal style, with specific terminology from the fields of Education and Psychology. As the texts are research articles, they also consist of a lot of data presented in tables. Such texts contain a lot of information, are very detailed, and have to be very precisely written. The language used has to be understandable and clear, but specific terminology may only be known to a specific group of readers.

The machine-translated texts were translated by onlinedoctranslator.com, where the texts were uploaded in a single document.

The next step was to upload the corpora to Sketch Engine (https://www.sketchengine.eu/) in order to obtain data regarding basic information about the texts, i.e. the number of tokens, words, and sentences, which is presented in the following table:

	SOURCE TEXTS	**HT TEXTS**	**MT TEXTS**
tokens	23 019	28 661	26 451
words	18 699	22 756	21 987
sentences	750	776	757

Table 1. Differences in the number of words between corpora

The texts that are translations done by a human translator have more words and sentences in total than the source texts and the machine-translated texts. This is presumably because human translators are translating the meaning and trying to convey it, not just words. Sometimes, it is necessary to use more words in one language than the other for certain phrases, because there are oftentimes no equivalents in the target language for phrases from the source language. Word order, reorganization and the grammatical structure of a sentence should also be taken into account. Also, the texts used for this research contain specific terminology which may not have a direct equivalent in English.

This might indicate that machines tend to simplify, and human translators might use strategies such as addition to make the target text more comprehensible. Machines may use word-for-word translation strategy, which refers to direct translation, while human translators may use sense-for-sense translation strategy more, trying to translate the meaning of the whole sentence or phrase.

Since nouns and noun phrases make up the majority of special field terminology, only noun phrases have been selected for further analysis.

Regarding the number of sentences, this study included 750 sentences in Croatian, 776 in texts translated into English by a human translator, and 757 in machine translated texts. This can be accounted for by semantic differences in Croatian and English and breaking down the sentences due to readability of the text in target language.

Using the *Terminology extraction option* in Sketch Engine, keywords were extracted from the corpus. The keywords have then been subjected to manual classification and filtering based on the following criteria:

1. Noise was eliminated from the corpus (e.g., doublets such as reading-reading were not included in the analysis)
2. Selection of phrases that might be ambiguous or difficult to interpret correctly.

Finally, 48 phrases were selected for the analysis. However, only 21 (44 %) were used for the survey. The reason for that is because some machine translated phrases were the same as the ones translated by a human translator, so it was decided not to include them in the research. Phrases that were selected for the analysis are presented in *Table 2*. Phrases that were actually used in the survey are in **bold**.

CROATIAN	MT	HT
1. daroviti učenik	gifted student	gifted student
2. lektira	**school reading**	**reading assignment**
3. učenik s teškoćama učenja	student with learning disabilities	student with learning disabilities
4. čitalačka pismenost	reading literacy	reading literacy
5. darovita djeca	gifted children	gifted children
6. teškoća učenja	**learning difficulty**	**learning disability**
7. nastava lektire	**teaching of reading**	**reading assignment class**
8. teorija odgoja	**theory of education**	**theory of upbringing**
9. daroviti učenik s ADHD-om	gifted student with ADHD	gifted student with ADHD
10. književni tekst	literary text	literary text
11. školsko postignuće	school achievement	school achievement
12. procjena darovitosti	assessment of giftedness	assessment of giftedness
13. nastava lektire	reading class	reading class
14. kognitivna sposobnost	cognitive ability	cognitive ability
15. lektirni naslov	reading title	reading assignment title
16. nedaroviti učenik	non-gifted student	non-gifted student
17. program za darovite	**program for gifted students**	**gifted program**
18. samostalno čitanje	independent reading	independent reading
19. književno djelo	literary work	literary work
20. teškoća učenja	learning disability	learning difficulty
21. obrazovna praksa	educational practice	educational practice
22. kritičko čitanje	critical reading	critical reading
23. razumijevanje pročitanoga	**comprehension of the read**	**reading comprehension**

24. temeljna obrazovna kompetencija	**fundamental educational competence**	**basic educational competence**
25. teškoća u verbalnoj obradi	difficulty in verbal processing	difficulty in verbal processing
26. motrenje razumijevanja pročitanoga	**observing the comprehension of the read**	**monitoring reading comprehension**
27. neatraktivni lektirni naslov	**unattractive reading title**	**unattractive reading assignment title**
28. učenik s teškoćama čitanja	**student with reading difficulties**	**student with reading disabilities**
29. lektirni naslov	**reading title**	**literary title**
30. čitalačka kompetencija	reading competency	reading competency
31. Okvir nacionalnog kurikuluma	**Framework of the National Curriculum**	**National curriculum framework**
32. pedagog	pedagogue	pedagogue
33. odgoj	**education**	**upbringing**
34. gimnazija	**high school**	**grammar school**
35. obrazovni sustav	educational system	educational system
36. konotativno čitanje	connotative reading	connotative reading
37. samoregulacija učenja	**self-regulation of learning**	**self-regulated learning**
38. strategija učenja	learning strategy	learning strategy
39. strategija poučavanja	teaching strategy	teaching strategy
40. učeničko zalaganje	**student efforts**	**students' commitment**
41. procjena učenika	**students' assessment**	**pupils' assessment**

42. odgojno-obrazovni ciljevi	**educational goals**	**educational aims**
43. ishodi	outcomes	outcomes
44. odgajatelj	educator	educator
45. odgojno-obrazovni plan	educational plan	education plan
46. dvostruko izuzetni učenici	**doubly exceptional students**	**twice-exceptional students**
47. završan razred	**final grade**	**final year**
48. roditeljski sastanak	**parent meeting**	**parent-teacher conference**

Table 2. Phrases selected for the analysis.

The research was conducted in the form of a survey via Google Forms (see Appendix 1). The survey was written in English. The first part of the survey is introductory and consists of the aim of the research, which was to evaluate whether a certain phrase was translated better by a machine or a human translator, and some other information, such as, that the original phrases are in Croatian and were translated into English, and that the survey is completely anonymous and confidential, as well as an e-mail address for contact, should the evaluators have any queries regarding the research.

The second part of the survey consists of collecting demographic data, i.e. age, sex, and highest educational degree received. I also wanted to find out whether the evaluators have a degree in English language or Translation, any experience in translating, and if they are a native English speaker. This was all done in a form of multiple-choice questions.

The main part of the survey depended on whether the evaluators chose that they are native English speakers or not (i.e., they are Croatian), because a separate survey was created for each.

The evaluators who are also native speakers of English were presented with 21 phrases. Each section offered two phrases in English, both of which were a translation of the same phrase in Croatian, except one was translated by a human translator, and one was a machine-translated

phrase. They had to select the phrase out of the pair that seems more appropriate to them as native English speakers. The order in which the phrases were presented was random, so there was no way for the evaluators to know which phrase was translated by a human, and which one was translated by a machine.

The evaluators who are not native English speakers (i.e., whose L1 is Croatian) were also presented with the same 21 phrases, but they were also given the source phrase in Croatian. However, they had to choose which phrase they thought was translated by a human translator and which was a machine-translated phrase.

All the questions were multiple-choice for both groups of evaluators, meaning that they were able to select one option out of the two given in each section. There were no open-ended questions or sections to add any additional information.

4.4.Evaluators

There were 40 evaluators taking part in this research, 19 (47.5%) female, 20 (50%) male, and 1 (2.5%) that preferred not to say. Three (3, i.e. 7.5%) evaluators selected that their age range is "18-20", 21 (52.5%) of them selected that their age range is "21-25", 5 (12.5%) chose that their age range is "26-30", 5 (12.5%) chose "31-40", 3 (7.5%) chose "41-50", and 3 (7.5%) evaluators selected that they are "≥51". Twenty (20, i.e. 50%) evaluators selected that their highest educational degree is a Bachelor's degree, 10 (25%) have a Master's degree, 5 (12.5%) have finished high school education, and 5 (12.5%) selected that they have a PhD. Ten (10, i.e. 25%) evaluators selected that they have a degree in English language or Translation. Nineteen (19, i.e. 47.5%) evaluators selected that they have experience in translating. Twenty (20, i.e. 50%) evaluators out of the 40 are native English speakers, while the other 20 (50%) are native Croatian speakers.

5. Results

This chapter presents the quantitative results of the survey, followed by a qualitative analysis of translation equivalents.

The first section of this chapter presents results from those evaluators who are not native English speakers.

The following table (*Table 3*) presents the results from the part of the survey that was created for non-native speakers. In the table, the source phrases in Croatian, the translations of the source phrase (machine translation = MT and human translation = HT), number of evaluators that selected that the phrase is translated by a machine, and the accuracy in percentages are presented. Accuracy refers to the percentage of evaluators that have correctly selected whether the phrase is a machine or a human translation of the original phrase. Phrases that are human-translated are in **bold.**

Original phrase (in Croatian)	**Translations of the original phrase (MT, HT)**	**Number of evaluators that chose that the phrase is MT**	**Accuracy**
lektira	school reading	13	65 %
	reading assignment	7	
teškoća učenja	learning difficulty	8	40 %
	learning disability	12	
nastava lektire	teaching of reading	14	70 %
	reading assignment class	6	
teorija odgoja	theory of education	8	40 %
	theory of upbringing	12	
program za darovite	program for gifted students	4	20 %
	gifted program	16	
razumijevanje pročitanoga	comprehension of the read	12	60 %
	reading comprehension	8	

temeljna obrazovna kompetencija	fundamental educational competence	12	60 %
	basic educational competence	8	
motrenje razumijevanja pročitanoga	observing the comprehension of the read	14	70 %
	monitoring reading comprehension	6	
neatraktivni lektirni naslov	unattractive reading title	12	60 %
	unattractive reading assignment title	8	
učenik s teškoćama čitanja	student with reading difficulties	10	50 %
	student with reading disabilities	10	
lektirni naslov	reading title	8	40 %
	literary title	12	
Okvir nacionalnog kurikuluma	Framework of the National Curriculum	12	60 %
	National Curriculum Framework	8	
odgoj	education	9	45 %
	upbringing	11	
gimnazija	high school	7	35 %
	grammar school	13	
samoregulacija učenja	self-regulation of learning	14	70 %
	self-regulated learning	6	
učeničko zalaganje	student efforts	9	45 %
	students' commitment	11	
procjena učenika	students' assessment	4	20 %

	pupils' assessment	16	
odgojno-obrazovni ciljevi	educational goals	13	65 %
	educational aims	7	
dvostruko izuzetni učenici	doubly exceptional students	9	45 %
	twice exceptional students	11	
završan razred	final grade	15	75 %
	final year	5	
roditeljski sastanak	parent meeting	13	65 %
	parent-teacher conference	7	

Table 3. Results of the survey for non-native speakers

Approximately 52.38% of the phrases were guessed correctly to be machine-translated phrases by the evaluators. The following formula was used to calculate the percentage.

$$\text{Percentage of correct guesses} = \left(\frac{\text{Total number of correct guesses}}{\text{Total number of phrases evaluated}}\right) \times 100$$

$$\text{Percentage of correct guesses} = \left(\frac{220}{420}\right) \times 100$$

$$\text{Percentage of correct guesses} = 52.38\%$$

Figure 1. Formula used to calculate the percentage for the overall result of the research

The following table (*Table 4*) presents the results of the survey for native speakers of English language. The first column presents the source phrases in Croatian, the second column presents the translations of the source phrases (MT and HT). The third column presents the number of evaluators that selected that a certain phrase is more acceptable than the other. Phrases that are human-translated are in **bold**.

Original phrase (in Croatian)	Translations of the original phrase (MT, HT)	Number of evaluators that chose that the phrase is more acceptable (max. 20)
lektira	school reading	4
	reading assignment	16
teškoća učenja	learning difficulty	14
	learning disability	6
nastava lektire	teaching of reading	7
	reading assignment class	13
teorija odgoja	theory of education	20
	theory of upbringing	0
program za darovite	program for gifted students	16
	gifted program	4
razumijevanje pročitanoga	comprehension of the read	0
	reading comprehension	20
temeljna obrazovna kompetencija	fundamental educational competence	7
	basic educational competence	13
motrenje razumijevanja pročitanoga	observing the comprehension of the read	0
	monitoring reading comprehension	20
neatraktivni lektirni naslov	unattractive reading title	7
	unattractive reading assignment title	13

učenik s teškoćama čitanja	student with reading difficulties	16
	student with reading disabilities	4
lektirni naslov	reading title	4
	literary title	16
Okvir nacionalnog kurikuluma	Framework of the National Curriculum	3
	National Curriculum Framework	17
odgoj	education	19
	upbringing	1
gimnazija	high school	18
	grammar school	2
samoregulacija učenja	self-regulation of learning	0
	self-regulated learning	20
učeničko zalaganje	student efforts	9
	students' commitment	11
procjena učenika	students' assessment	16
	pupils' assessment	4
odgojno-obrazovni ciljevi	educational goals	16
	educational aims	4
dvostruko izuzetni učenici	doubly exceptional students	13
	twice-exceptional students	7
završan razred	final grade	4
	final year	16
	parent meeting	4

roditeljski sastanak	**parent-teacher conference**	16

Table 4. Results of the survey for native speakers of English

It should be noted that human translation is taken as gold standard for both native and non-native speakers in this survey.

6. Discussion

6.1.Non-native speakers' survey

For the Croatian phrase *lektira*, 13 evaluators guessed correctly that the phrase *school reading* is the machine translation, while the other 7 evaluators thought that the phrase *reading assignment* is the machine translation, even though the latter is a human translation. While the evaluators were rather certain in recognizing the machine-translated phrase in this instance (with a 65% success rate) this was not always the case. For example, 12 evaluators selected that the phrase *learning disability* is the machine translation of the Croatian phrase *teškoća učenja*, while the other 8 evaluators selected the phrase *learning difficulty* as the machine translation. As can be seen from these two examples, the words *difficulty* and *disability* are much more abstract and are 'closer' synonyms, hence, they were confusing for the evaluators.

For the Croatian phrase *nastava lektire*, 14 evaluators (70%) guessed correctly that the phrase *teaching of reading* is the machine translation. The other 6 evaluators (30%) selected the human-translated phrase *reading assignment class* as the machine translation. For the phrase *teorija odgoja*, 8 evaluators (40%) guessed correctly that *theory of education* is the machine translation, while 12 evaluators (60%) selected the human-translated phrase *theory of upbringing* as machine translation. Regarding the Croatian phrase *program za darovite*, 16 evaluators (80%) selected that the human-translated phrase *gifted program* is the machine translation, while 4 evaluators (20%) guessed correctly that the phrase *selected program for gifted students* is the machine translation. 8 evaluators (40%) selected that the human-translated phrase *reading comprehension* is the machine translation of the phrase *razumijevanje pročitanoga.* The other 12 evaluators (60%) guessed correctly that the phrase *comprehension of the read* is the machine translation. For the Croatian phrase *temeljna obrazovna kompetencija*, 12 evaluators (60%) guessed correctly that the phrase *fundamental educational competence* is the machine translation, while 8 evaluators (40%) selected the human-translated phrase *basic educational competence* as the machine translation. Regarding the Croatian phrase *motrenje razumijevanja pročitanoga, observing the comprehension of the read* was guessed correctly by 14 evaluators (70%) as the machine translation, while 6 evaluators (30%) chose the human-translated phrase *monitoring reading comprehension* as the machine translation. Regarding the Croatian phrase *neatraktivni*

lektirni naslov, 8 evaluators (40%) chose the human-translated phrase *unattractive reading assignment title* as the machine-translated phrase, and 12 evaluators (60%) guessed correctly that the phrase *unattractive reading title* is the machine translation. For the Croatian phrase *učenik s teškoćama čitanja*, 10 evaluators (50%) guessed correctly that the phrase *student with reading difficulties i*s the machine translation, while the other 10 evaluators (50%) selected the human-translated phrase *student with reading disabilities* as the machine translation. The phrase *reading title* was guessed correctly by 8 evaluators (40%) as the machine translation for the original phrase *lektirni naslov*, while the other 12 evaluators (60%) selected the human-translated phrase *literary title* as the machine translation. For the Croatian phrase *Okvir nacionalnog kurikuluma,* 12 evaluators (60%) guessed correctly that the phrase *Framework of the National Curriculum i*s the machine translation, while the other 8 evaluators (40%) selected the human-translated phrase *National Curriculum Framework.* Regarding the Croatian phrase *odgoj*, 9 evaluators (45%) guessed correctly that *education* is the machine translation, while 11 evaluators (55%) chose the human-translated phrase *upbringing* as the machine translation. For the original phrase *gimnazija*, 7 evaluators (35%) guessed correctly that the phrase *high school* is the machine translation, and 13 evaluators (65%) chose the human-translated phrase *grammar school* as the machine translation. 6 evaluators (30%) selected the human-translated phrase *self-regulated learning* as the machine translation of the original phrase *samoregulacija učenja,* while the other 14 evaluators (70%) guessed correctly that *self-regulation of learning* is the machine translation. For the original phrase *učeničko zalaganje,* 9 evaluators (45%) guessed correctly that the phrase *student efforts* is the machine translation, while the other 11 evaluators (55%) selected the human-translated phrase *students' commitment* as the machine translation. Regarding the Croatian phrase *procjena učenika,* 16 evaluators (80%) chose that the human-translated phrase *pupils' assessment* is the machine translation, while 4 evaluators (20%) guessed correctly that the phrase *students' assessment* is the machine translation. 13 evaluators (65%) guessed correctly that the phrase *educational goals* is the machine translation of the original phrase *odgojno-obrazovni ciljevi,* while 7 evaluators (35%) chose the human-translated phrase *educational aims* as the machine translation. Regarding the original phrase *dvostruko izuzetni učenici,* 11 evaluators (55%) chose that the human-translated phrase *twice-exceptional students* is the machine translation of the original phrase, while 9 evaluators (45%) guessed correctly that the phrase *doubly exceptional students* is the machine translation. For the original phrase *završan razred,* 15 evaluators (75%) guessed correctly that the

phrase *final grade* is the machine translation, and 5 evaluators (25%) selected the human-translated phrase *final year* as the machine translation. Regarding the Croatian phrase *roditeljski sastanak,* 7 evaluators (35%) chose that the human-translated phrase *parent-teacher conference* is the machine translation, while 13 evaluators (65%) guessed correctly that the phrase *parent meeting* is the machine translation.

The lowest accuracy (20%) is in two phrases, *program za darovite* and *procjena učenika*, while the highest accuracy is for the phrase *završan razred* (75%). Overall, more than half of the non-native speaking evaluators (approximately 52.38%) were able to identify which phrase is the machine translation. It can be concluded that less than half of the non-native speaking evaluators (approximately 47.6%) were not able to identify which phrase is a machine translation, and that more than half of the non-native evaluators are able to distinguish between machine-translated and human-translated phrases in English.

6.2.Native speakers' survey

16 evaluators (80%) found the human-translated phrase *reading assignment* more acceptable than the phrase *school reading.* 14 evaluators (70%) selected the machine-translated phrase *learning difficulty* as more acceptable than the phrase *learning disability*. 13 evaluators (65%) selected the human-translated phrase *reading assignment class* as more acceptable than the phrase *teaching of reading.* 20 evaluators (100%) chose the machine-translated phrase *theory of education* as more acceptable than the phrase *theory of upbringing.* While *theory of upbringing* is actually a human translation, it is important to note that native speakers of English find the machine translation (*theory of education*) more acceptable. According to the Cambridge Dictionary (2024), *education* refers to "the process of teaching or learning, especially in a school or college (...)", while *upbringing* refers to "the way in which you are treated and educated when young, especially by your parents (...)". However, in the context of Croatia and its education system, we can find that education usually falls under the term of upbringing, as we understand the term education as *odgoj i obrazovanje.* Here we can conclude that it can be seen as a culturally marked concept. 16 evaluators (80%) find the machine-translated phrase *program for gifted students* more acceptable than the phrase *gifted program.* 20 evaluators (100%) selected the human-translated phrase *reading comprehension* as more acceptable than the phrase *comprehension of the read.* 13

evaluators (65%) selected that the human-translated phrase *basic educational competence* is more acceptable than the phrase *fundamental educational competence.* 20 evaluators (100%) selected that the human-translated phrase *monitoring reading comprehension* is more acceptable than the phrase *observing the comprehension of the read.* 13 evaluators (65%) chose the human-translated phrase *unattractive reading assignment title* as more acceptable than *unattractive reading title.* 16 evaluators (80%) selected the machine-translated phrase *student with reading difficulties* as more acceptable than the phrase *student with reading disabilities.* 16 evaluators (80%) selected the human-translated phrase *literary title* as more acceptable than the phrase *reading title.* 17 evaluators (85%) selected that the human-translated phrase *National Curriculum Framework* is more acceptable than the phrase *Framework of the National Curriculum.* 19 evaluators (95%) found the machine-translated phrase *education* more acceptable than *upbringing.* 18 evaluators (90%) selected the machine-translated phrase *high school* as more acceptable than the phrase *grammar school.* It is understandable how this might create confusion and ambiguity, as phrases *high school* and *grammar school* essentially refer to the same thing, however, it does depend on the English speaking are. 20 evaluators (100%) chose the human-translated phrase *self-regulated learning* as more acceptable than *self-regulation of learning.* 11 evaluators (55%) selected the human-translated phrase *students' commitment* as more acceptable than the phrase *student efforts.* 16 evaluators (80%) found the machine-translated phrase *students' assessment* more acceptable than *pupils' assessment.* 16 evaluators (80%) chose the machine-translated phrase *educational goals* more acceptable than *educational aims.* 13 evaluators (65%) selected the machine-translated phrase *doubly exceptional students* as more acceptable than *twice-exceptional students.* 16 evaluators (80%) selected that the human-translated phrase *final year* is more acceptable than *final grade.* 16 evaluators (80%) found the human-translated phrase *parent-teacher conference* more acceptable than the phrase *parent meeting.* It is important to note that these phrases could have caused some confusion with the evaluators because they do not refer to the exactly same thing. According to Collins Dictionary (2024), *parent-teacher meeting* refers to "an occasion when the parents of children at a school and their teachers come together (...) in order to discuss the progress or work of the children", while the term *parent meeting* could refer to a meeting between the parents of the students, with no presence of the teachers. If we refer to the source phrase (*roditeljski sastanak*), in my opinion, its meaning best fits the translation *parent-teacher conference.*

In some instances, native-speaking evaluators have confidently identified a phrase as either a machine translation or a human translation, with an accuracy rate of 100%. For example, for the original phrase *teorija odgoja*, all 20 evaluators (100%) have agreed that they find the machine-translated phrase *theory of education* more acceptable than the human translation. However, for phrases *razumijevanje pročitanoga, samoregulacija učenja* and *motrenje razumijevanja pročitanoga,* all 20 evaluators selected in each instance that the human translations of the original phrases are more acceptable than the machine translations.

Overall, less than half of the evaluators (approximately 46%) selected the machine-translated phrases as more acceptable than the human translations, i.e., approximately 54% of the evaluators selected human-translated phrases as more acceptable than machine translations. It can be concluded here that native speakers find human translations more acceptable than phrases translated by a machine.

7. Conclusion

Based on the assumption that machine translation is improving rapidly, the main aim of this thesis was to contribute to assessing the quality of machine translation for the language pair Croatian-English. The evaluation of human versus machine translation yielded a varied perspective and different levels of acceptability were found for each approach among the evaluators. It is important to note that the research was done using a small sample size.

Approximately 52.38% of non-native evaluators guessed correctly which phrases are machine-translated, i.e. which phrases are a human translation. Overall, more than half of the non-native evaluators were able to distinguish between machine-translated and human-translated phrases in English.

Approximately 46% of the native English-speaking evaluators found the machine-translated phrases more acceptable than the ones translated by a human translator, i.e. approximately 54% of evaluators selected human translations as more acceptable than the machine-translated phrases. Overall, more than half of the native evaluators find human-translated phrases more acceptable than phrases translated by a machine.

As for the research questions outlined in section 4.2 the following answers can be provided:

Question 1: Are native speakers of English able to distinguish between machine-translated and human-translated phrases in English? Do they perceive machine-translated phrases as acceptable as the human-translated phrases?

Interestingly, approximately 46% of evaluators preferred machine-translated phrases, while approximately 54% found human-translated ones more acceptable. Due to the very low difference in results between selecting whether native speakers of English find the machine translation more acceptable than the human translation, it is difficult to answer this question. It can be concluded that, even though more than half of them (54%) selected human-translated phrases as more acceptable, they still found machine translation more acceptable in a lot of instances. Therefore, they do perceive machine-translated phrases almost as acceptable as the human-translated phrases.

Question 2: Are non-native speakers of English able to distinguish between machine-translated and human-translated phrases in English?

It can be concluded that non-native speakers of English are able to distinguish between machine-translated and human-translated phrases in English. However, similar to the findings of native English-speaking evaluators, the difference is very slight.

Question 3: Given the improvement of machine translation systems, is the boundary between natural (human) and machine-translated language becoming less clear? OR is it becoming increasingly difficult to distinguish machine-translated phrases (texts) from human produced translations?

From the results of this research, it can be concluded that the boundary between natural and machine-translated language is becoming less clear. According to the findings, it is becoming more difficult to distinguish machine-translated phrases and texts from human produced translations.

These findings may suggest that machine translation has advanced significantly in producing acceptable translations, particularly in certain contexts or for specific types of content. However, the preference for human-translated phrases among a slight majority of evaluators emphasizes the continued value of human expertise, especially in idiomatic expressions, cultural references, etc.

Furthermore, the disparities in preferences may also reflect individual variations in tolerance for errors, stylistic preferences, or familiarity with machine translation technology. Therefore, it may not be appropriate to use a one-size-fits-all approach to evaluate translations. Instead, it is essential to have a nuanced understanding of the strengths and limitations of both human and machine translation.

In conclusion, although machine translation has potential and shows efficiency, human translation remains indispensable in delivering high-quality, contextually appropriate translations. Future research should explore more the understanding of the specific contexts and criteria under which each method excels, thereby improving the effectiveness of both human and machine translation practices. The research should be done with a larger sample size, and the factor of knowledge of English and having experience in translating should be taken into account, as it might be relevant for the results.

Reference List

Ahrenberg, L. (2017). Comparing Machine Translation and Human Translation: A Case Study. *Proceedings of the Workshop Human-Informed Translation and Interpreting Technology,* 21-28. DOI: 10.26615/978-954-452-042-7_003.

Babych, B. and Hartley, A.F. (2008). Sensitivity of Automated MT Evaluation Metrics on Higher Quality MT Output: BLEU vs Task-Based Evaluation Methods. *International Conference on Language Resources and Evaluation,* 2133-2136.

Brust Nemet, M. and Vrdoljak, G. (2022). Gender (In)equality in Child-Rearing and Housework between Mothers and Fathers – Children's Perspective. *Croatian Journal of Education, 24*. (2.), 429-455. DOI: https://doi.org/10.15516/cje.v24i2.4554

Cambridge Dictionary. (2024). Education. In cambridge.dictionary,org. Retrieved February 1, 2024, from https://dictionary.cambridge.org/dictionary/english/education.

Collins Dictionary. (2024). Parent-teacher meeting. In collinsdictionary.com. Retrieved February 1, 2024, from https://www.collinsdictionary.com/dictionary/english/parent-teacher-meeting

Craciunescu, O., Gerding-Salas, C. and Stringer-O"Keeffe, S. (2007). Machine Translation and Computer-Assisted Translation : A New Way of Translating ? et al. (2007). "Machine Translation and Computer-Assisted Translation: A New Way of Translating?" *Translation Journal* 8.

Cvitković, D. and Stošić, J. (2022). Gifted Students with Disabilities. *Croatian Journal of Education, 24.* (3.), 949-986. DOI: https://doi.org/10.15516/cje.v24i3.4470

Dorr, B., Snover, M. and Madnani, N. (2011). Chapter 5.1 introduction. In Olive, J., McCary, J. and Christianson, C. (eds), Handbook of Natural Language Processing and Machine Translation. DARPA Global Autonomous Language Exploitation. New York: Springer, pp. 801–803.

Dunđer, I. (2020). Machine Translation System for the Industry Domain and Croatian Language. *Journal of Information and Organizational Sciences, 44*(1), 33-50. DOI: https://doi.org/10.31341/jios.44.1.2

Escribe, M. (2019). Human Evaluation of Neural Machine Translation: The Case of *Deep Learning*. *Proceedings of the Second Workshop Human-Informed Translation and Interpreting Technology associated with RANLP 2019,* 36-46. DOI: 10.26615/issn.2683-0078.2019_005.

Han, B. (2020). Translation, from Pen-and-Paper to Computer-Assisted Tools (CAT Tools) and Machine Translation (MT). *Proceedings,* 63. 56. DOI: 10.3390/proceedings2020063056.

Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., Koo, S., and Lim, H. (2023). A Survey on Evaluation Metrics for Machine Translation. *Mathematics* 11(4):1006. DOI: https://doi.org/10.3390/math11041006

Licht, D., Gao, C., Lam, J., Guzmán, F., Diab, M.T. and Koehn, P. (2022). Consistent Human Evaluation of Machine Translation across Language Pairs. Conference of the Association for Machine Translation in the Americas. (2022). Consistent Human Evaluation of Machine Translation across Language Pairs. *In Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track),* 309-321. DOI: 10.48550/arXiv.2205.08533.

Lin, Chin-Yew and Och, F. (2004). ORANGE: a Method for Evaluating Automatic Evaluation Metrics for Machine Translation. *COLING '04: Proceedings of the 20th international conference on Computational Linguistics,* 501-507. DOI: https://doi.org/10.3115/1220355.1220427

Machine Translation 101 – Part 3. (2021, January 5). Defined.ai. https://resources.defined.ai/blog/machine-translation-101-part-3/

Maučec, M.S., and Donaj, G. (2019). Machine Translation and the Evaluation of Its Quality. *Recent Trends in Computational Intelligence.* DOI: 10.5772/intechopen.89063.

Munkova, D., Munk, M., Benko, L. and Stastny, J. (2021). MT Evaluation in the Context of Language Complexity. *Complexity. 2021*, 1-15. DOI: https://doi.org/10.1155/2021/2806108

Okpor, M.D. (2014). Machine Translation Approaches: Issues and Challenges. *International Journal of Computer Science Issues,* 11: 158-165.

Online Doc Translator. (n.d.). Accessed 1 June, 2024 on https://www.onlinedoctranslator.com/en/

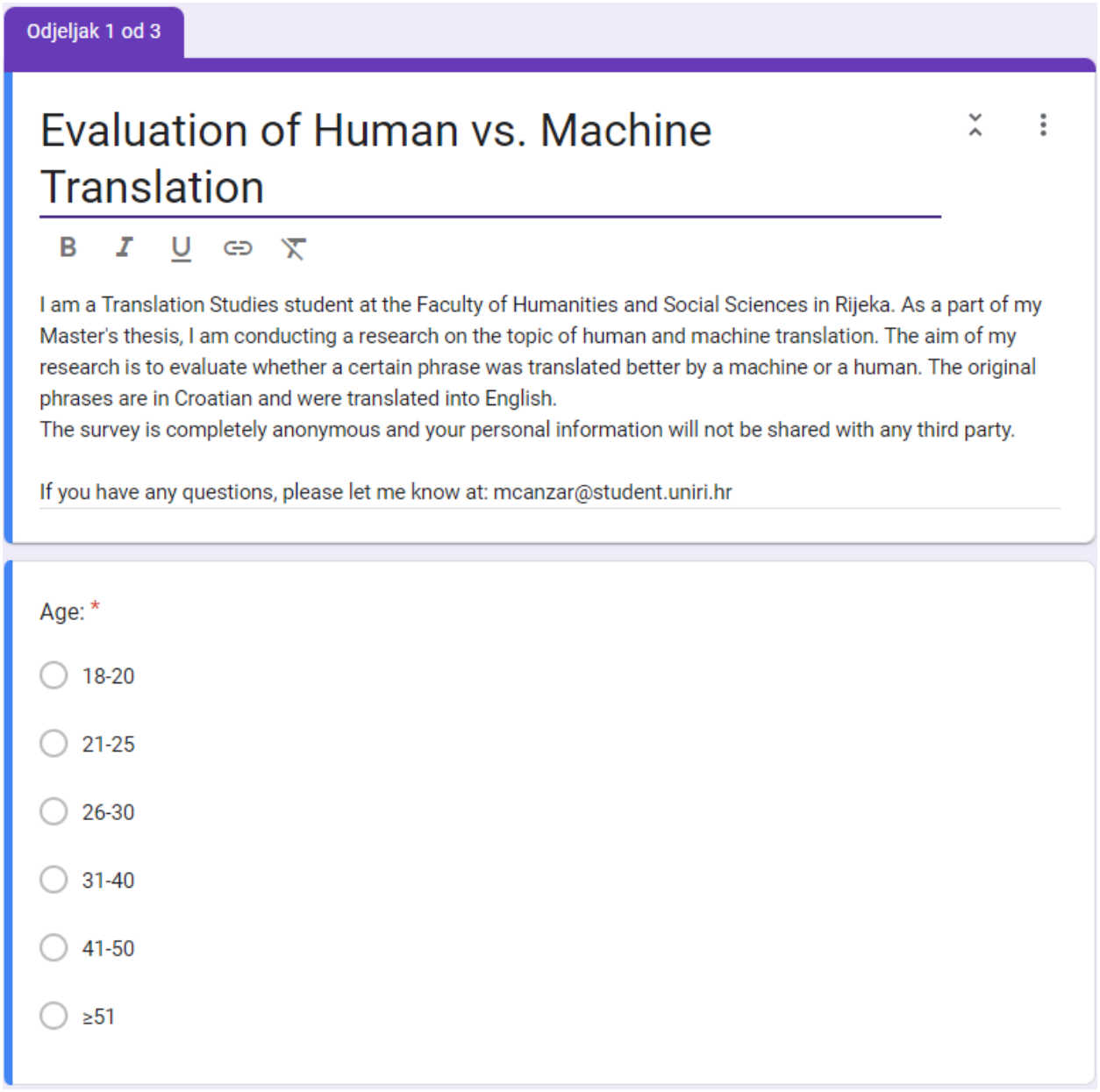
Pavičić Vukičević, J., Prpić, M. and Cajner Mraović, I. (2022). Students' Attitude towards Reading Assignments and Reading: The Perception of Students and Teachers in Secondary Schools in Zagreb. *Croatian Journal of Education, 24.* (1.), 127-160. DOI: https://doi.org/10.15516/cje.v24i1.4282

Sanders, G., Przybocki, M., Madnani, N. and Snove, M. (2010). Part 5: Machine Translation Evaluation Chapter 5.1.2. Human Subjective Judgments.

Sketch Engine. (n.d.). Accessed 15 June 2024 on https://www.sketchengine.eu/.

Šimunović, Z., Vidulin, S. and Miljković, D. (2022). Flow Experiences in Adolescents: Comparison of Musically Educated, Athletically Active, and Other Adolescents. *Croatian Journal of Education, 24.* (4.), 1205-1227. DOI: https://doi.org/10.15516/cje.v24i4.4703

Appendix 1



Odjeljak 1 od 3

Evaluation of Human vs. Machine Translation

I am a Translation Studies student at the Faculty of Humanities and Social Sciences in Rijeka. As a part of my Master's thesis, I am conducting a research on the topic of human and machine translation. The aim of my research is to evaluate whether a certain phrase was translated better by a machine or a human. The original phrases are in Croatian and were translated into English.
The survey is completely anonymous and your personal information will not be shared with any third party.

If you have any questions, please let me know at: mcanzar@student.uniri.hr

Age: *

- 18-20
- 21-25
- 26-30
- 31-40
- 41-50
- ≥51

Sex: *

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other

Highest educational degree received: *

- ○ Primary education
- ○ Highschool education
- ○ Bachelor's degree
- ○ Master's degree
- ○ PhD

Do you have a degree in English language or Translation? *

- ○ Yes
- ○ No

Do you have any experience in translating? *

- Yes
- No

Are you a native English speaker? *

- Yes
- No

Odjeljak 2 od 3

Human vs. Machine Translation

In the following questions, you will be given phrases in Croatian and their translations into English. You should select which translation you think was done by a human and which by a machine. Choose MT for machine translation and HT for human translation. Please note that, in each example, you cannot choose the same option for both translations.

lektira *

	MT	HT
school reading	○	○
reading assignment	○	○

teškoća učenja *

	MT	HT
learning disability	○	○
learning difficulty	○	○

nastava lektire *

	MT	HT
teaching of reading	○	○
reading assignment class	○	○

teorija odgoja *

	MT	HT
theory of education	○	○
theory of upbringing	○	○

program za darovite *

	MT	HT
gifted program	○	○
program for gifted students	○	○

razumijevanje pročitanoga *

	MT	HT
reading comprehension	○	○
comprehension of the read	○	○

temeljna obrazovna kompetencija *

	MT	HT
fundamental educational compet...	○	○
basic educational competence	○	○

motrenje razumijevanja pročitanoga *

	MT	HT
observing the comprehension of ...	○	○
monitoring reading comprehension	○	○

neatraktivni lektirni naslov *

	MT	HT
unattractive reading assignment t...	○	○
unattractive reading title	○	○

učenik s teškoćama čitanja *

	MT	HT
student with reading difficulties	○	○
student with reading disabilities	○	○

lektirni naslov *

	MT	HT
reading title	○	○
literary title	○	○

Okvir nacionalnog kurikuluma *

	MT	HT
Framework of the National Curric...	○	○
National Curriculum Framework	○	○

odgoj *

	MT	HT
education	○	○
upbringing	○	○

gimnazija *

	MT	HT
high school	○	○
grammar school	○	○

samoregulacija učenja *

	MT	HT
self-regulated learning	○	○
self-regulation of learning	○	○

učeničko zalaganje *

	MT	HT
student efforts	○	○
students' commitment	○	○

procjena učenika *

	MT	HT
pupils' assessment	○	○
students' assessment	○	○

odgojno-obrazovni ciljevi *

	MT	HT
educational goals	○	○
educational aims	○	○

dvostruko izuzetni učenici *

	MT	HT
twice-exceptional students	○	○
doubly exceptional students	○	○

završan razred *

	MT	HT
final grade	○	○
final year	○	○

roditeljski sastanak *

	MT	HT
parent-teacher conference	○	○
parent meeting	○	○

Odjeljak 3 od 3

Human vs. Machine Translation

You will be given two phrases which are translation equivalents of a Croatian phrase. One phrase was translated by a human translator and the other by a machine. Please select the phrase that you, as a native speaker of English, find more acceptable.

Please select the phrase that you find more acceptable. *

- ○ school reading
- ○ reading assignment

Please select the phrase that you find more acceptable. *

- ○ learning difficulty
- ○ learning disability

Please select the phrase that you find more acceptable. *

- teaching of reading
- reading assignment class

Please select the phrase that you find more acceptable. *

- theory of education
- theory of upbringing

Please select the phrase that you find more acceptable. *

- gifted program
- program for gifted students

Please select the phrase that you find more acceptable. *

- comprehension of the read
- reading comprehension

Please select the phrase that you find more acceptable. *

- fundamental educational competence
- basic educational competence

Please select the phrase that you find more acceptable. *

- observing the comprehension of the read
- monitoring reading comprehension

Please select the phrase that you find more acceptable. *

- unattractive reading title
- unattractive reading assignment title

Please select the phrase that you find more acceptable. *

- student with reading difficulties
- student with reading disabilities

Please select the phrase that you find more acceptable. *

- ○ reading title
- ○ literary title

Please select the phrase that you find more acceptable. *

- ○ Framework of the National Curriculum
- ○ National Curriculum Framework

Please select the phrase that you find more acceptable. *

- ○ self-regulation of learning
- ○ self-regulated learning

Please select the phrase that you find more acceptable. *

- ○ student efforts
- ○ students' commitment

Please select the phrase that you find more acceptable. *

- students' assessment
- pupils' assessment

Please select the phrase that you find more acceptable. *

- educational goals
- educational aims

Please select the phrase that you find more acceptable. *

- doubly exceptional students
- twice-exceptional students

Please select the phrase that you find more acceptable. *

- final grade
- final year

Please select the phrase that you find more acceptable. *

- parent meeting
- parent-teacher conference