Comparative Analysis of MemoQ Adaptive Generative Translation (AGT) with Conventional Machine Translation (MT) Engines

Aladrović, Josip

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Diplomski studij engleskog jezika i književnosti – prevoditeljski smjer i hrvatskog jezika i književnosti – nastavnički smjer

Josip Aladrović

Usporedna analiza adaptivnog generativnog prijevoda (AGT) s konvencionalnim strojnim prijevodom (MT)

Diplomski rad

Mentor: prof. dr. sc. Marija Omazić

Osijek 2024.

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Odsjek za engleski jezik i književnost

Diplomski studij engleskog jezika i književnosti – prevoditeljski smjer i hrvatskog jezika i književnosti – nastavnički smjer

Josip Aladrović

Usporedna analiza adaptivnog generativnog prijevoda (AGT) s konvencionalnim strojnim prijevodom (MT)

Diplomski rad

Znanstveno područje: Humanističke znanosti

Znanstveno polje: Filologija

Znanstvena grana: Anglistika

Mentor: prof. dr. sc. Marija Omazić

Osijek 2024.

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences of Osijek

Department of English Language and Literature

Double Major MA Study Programme in English Language and Literature – Translation and Interpreting Studies and Croatian Language and Literature – Education Studies

Josip Aladrović

Comparative Analysis of MemoQ Adaptive Generative Translation (AGT) with Conventional Machine Translation (MT) Engines

Master's Thesis

Supervisor: Dr. Marija Omazić, Full Professor

Osijek 2024

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences of Osijek

Department of English Language and Literature

Double Major MA Study Programme in English Language and Literature – Translation and Interpreting Studies and Croatian Language and Literature – Education Studies

Josip Aladrović

Comparative Analysis of MemoQ Adaptive Generative Translation (AGT) with Conventional Machine Translation (MT) Engines

Master's Thesis

Scientific area: Humanities

Scientific field: Philology

Scientific branch: English studies

Supervisor: Dr. Marija Omazić, Full Professor

Osijek 2024

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, 2024.

Josip Abdrović , 0165076709

Ime i prezime studenta, JMBAG

Abstract

This thesis aims to conduct a comprehensive comparative analysis between memoQ Adaptive Generative Translation (AGT) and conventional Machine Translation (MT) engines regarding their efficacy and accuracy in translating English language texts into Croatian. The study will investigate various aspects such as translation accuracy, fluency, and terminological consistency by the means qualitative analyses and of automatic evaluation provided by the MATEO evaluation platform and selected metrics (BLEU, ChrF, and COMET). Additionally, it will explore potential drawbacks associated with each translation approach. By employing empirical research methods and a case study, this thesis seeks to provide insights into the strengths and limitations of memoQ AGT compared to standard MT engines in meeting translation requirements in the English-Croatian language pair. The findings of this research are expected to contribute to the ongoing discourse on the advancements in machine translation technology.

Key words: memoQ AGT, machine translation, AI, MT engines, NMT, adaptive translation

Sažetak

Cilj je ovog rada sveobuhvatna usporedna analiza alata memoQ Adaptive Generative Translation (AGT) s ostalim konvencionalnim alatima za strojno prevođenje, s obzirom na učinkovitost i točnost prijevoda teksta s engleskog jezika na hrvatski jezik. Rad će istražiti razne aspekte poput točnosti prijevoda, tečnosti i terminološke dosljednosti u vidu kvalitativne analize i automatske procjene koju pruža platforma MATEO odabranim metrikama (BLEU, ChrF i COMET). Rad će također istražiti moguće nedostatke s obzirom na pojedini pristup prijevodu. Provedbom metoda empirijskog istraživanja i studije slučaja, ovaj rad cilja dati uvid u prednosti i ograničenja memoQ AGT-a u usporedbi sa standardnim alatima za strojno prevođenje s obzirom na zadovoljavanje kriterija prijevoda u jezičnoj kombinaciji engleski-hrvatski. Očekuje se da će rezultati ovog istraživanja doprinijeti tekućoj raspravi o napretku tehnologija strojnog prevođenja.

Ključne riječi: *memoQ AGT, strojni prijevod, UI, alati za strojno prevođenje, NMT adaptivno prevođenje*

Table of Contents

1. Introduction

In a world where technology is rapidly evolving to the point one can almost expect something new every single day, it is a bit out of the ordinary to wait a decade for the next breakthrough in any given field. When taking a closer look into the translation industry, we can clearly draw a conclusion that it is heavily dominated by NMT (neural machine translation), ever since it was first made available for public use. Since then, the only 'new' technology that had, or maybe still has, the opportunity to surpass the success and widespread use of NMT is adaptive machine translation. Unfortunately, adaptive machine translation has not seen the same kind of commercial success even though there are some studies and opinions on its better overall performance regarding terminological accuracy, consistency, tone and style. To name a few, Hanna Martikainen's "Ghosts in the machine: Can adaptive MT help reclaim a place for the human in the loop?" (2022), Yasmin Moslem's et al. "Adaptive Machine Translation with Large Language Models" (2023), Kirti Vashee's "Understanding Adaptive Machine Translation" (2023) written for ModernMT and so on.

Previous sentiment might change depending on the success and quality of memoQ AGT. In that regard, this paper aims to explore the possibilities, effectiveness and implications of using AI in translation paired with adaptive machine translation. By doing so, it will try to answer the following questions: Is there a noticeable quality increase when comparing memoQ AGT to conventional MT engines in regard to translation quality metrics? How different is memoQ AGTs' output compared to the baseline output of Microsoft Translator? Is memoQ AGT better at utilizing existing translation memories and reference documents as opposed to other NMT engines? How accurate is memoQ AGTs' output when compared to standard NMT systems and human translation regarding terminology, consistency, tone and style when translating administrative text that contains roughly 25 000 characters from English to Croatian. This will be done by comparing memoQ AGT with conventional MT engines such as Microsoft Translator, Google Translate and ModernMT.

Just by observing the image below, we can see how rapidly the field of AI translation has been evolving over the past couple of years with new models popping up left and right and the technology constantly evolving:

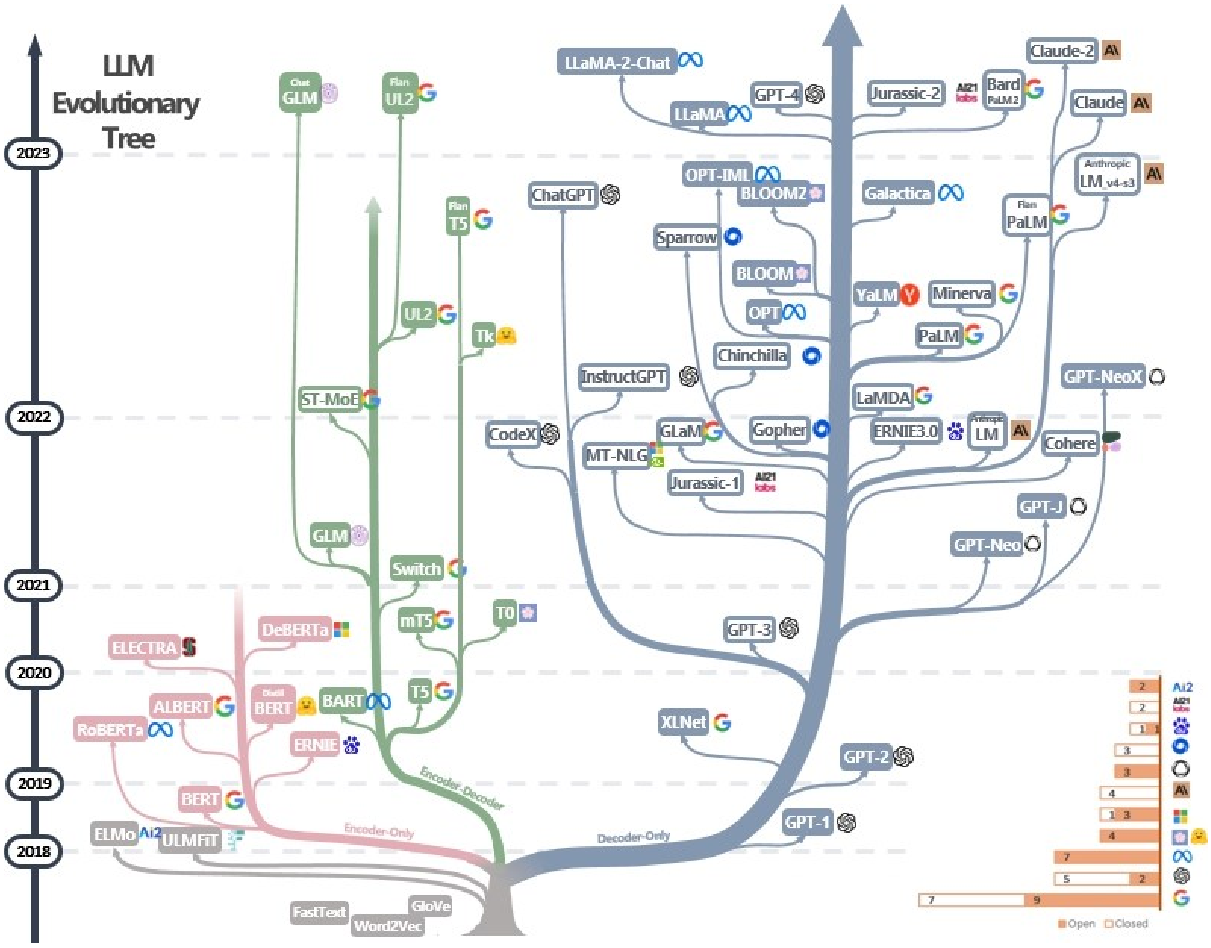

Figure 1: The Practical Guides for Large Language Models

This is a promising fact for AI in translation since the usefulness of it is, by now, well recognised. Provided that memoQ AGT, or maybe some other similar engines yet to be announced, can work with any LLM (large language model), we can only expect more improvements as AI technology evolves which may be a driving factor in making adaptive machine translation that makes use of AI technology more accurate, reliable and widespread.

If this paper achieves them aim set for itself earlier in the introduction, with the release of memoQ AGT, a new state-of-the-art translation tool, the world of translation might just be seeing the next step forward in the ever-improving translation industry. This paper will compare the results of three different MT engines against the AGT to ascertain whether adaptive machine translation powered by AI truly has the upper hand over conventional NMT.

2. Historical and theoretical overview

Machine translation (MT) has now existed for nearly a century, slowly progressing and always tackling what is thought to be one of the biggest tasks in its branch – computing and AI. Since the launch of NMT, MT has had its most significant jump in quality and overall public usage. While NMT and its usefulness and success are recognised worldwide, as previously mentioned, the same cannot be said for currently available tools that provide adaptive machine translation. However, it is important to distinguish between two different types of adaptation utilised by engines that will be covered and put to use in this research. One of the engines that will be mentioned later is ModernMT. It is an adaptive machine translation engine that learns and adapts itself by post-editing. On the other hand, memoQ AGT is a new engine that utilises AI technology and adapts itself during the translation process with available data from translator's local translation memories, termbases, and aligned or non-aligned corpora in memoQ LiveDocs as well as by communicating with the LLM. (*Understanding Adaptive Machine Translation)*

The beginning of machine translation dates back to the 1960s and 1970s, when the rule-based machine translation (RBMT) was invented. In an article about NMT for *Weglot*, Merve Alsan states that RMBT software translated the source text by analysing it word for word and then deciding how to translate each word by checking the ruleset developed by linguists in charge. (*A Quick Overview of Neural Machine Translation)* As you might expect, this type of translation did not provide the most accurate results, but it was a valuable start nonetheless.

After that, in the 1980s, came example-based machine translation (EBMT). In their paper titled "Example-Based Machine Translation with a Multi-Sentence Construction Transformer Architecture", Xiao, Zhou and Lepage defined EBMT in the following terms: "EBMT extracts knowledge from a corpus in two languages to perform translation. Concretely, the process of EBMT by analogy involves extracting analogical relationships in the source language to find the corresponding sentences in the target language and solve a sentence analogy." (2023: 72).

The 1990s then stepped forward with statistical machine translation (SMT), which is defined as a "...machine translation approach that uses large volumes of bilingual data to find the most probable translation for a given input." (*Statistical machine translation*) Statistical machine translation systems learn to translate by analysing the statistical relationships between original texts and their existing human translations." (*Statistical Machine Translation*), and its two most important components are the language model and the translation models. According to Kenny (2020: 163) in SMT, word sequences from a parallel bilingual corpus are aligned which forms a basis for the translation model. She also writes that it all started when "...researchers began to explore the possibility of exploiting corpora of already translated texts for automatic translation." Considering this fact, we can rightfully assume that the volume of provided data is crucial. This will be true not just for SMT but for all other machine translation systems to come.

It took a long time before a new breakthrough happened. Luckily, in 2015 NMT was invented, which was the first machine output that sounded human-like, fluid, and natural. NMT became available for public use the following year. To this day, various uses and forms of NMT remain the best option one can ask for when tackling translation challenges. With the 10-year anniversary just around the corner, the question arises: what will be the next breakthrough? On the other hand, the upsides are that it excels in morphology and syntax and does well with languages that are similar to one another. After being the dominant machine translation system for around three decades, SMT took over and replaced EBMT. It was dominant because it could automatically produce a system straight from the TM database, unlike EMBT systems, which, in return, made it more economically efficient. NMT was the next big breakthrough since it leveraged the benefits of neural networks and deep learning. The question that was hinted at the beginning of the paper is: Will AGT and similar adaptive machine translation engines that utilize AI surpass NMT and in what way?

3. Machine translation

In her chapter on *Technology in Translator Training,* Kenny states that "...the first machine translation (MT) systems went into operation in the 1960s but it was not until the 1980s that translator training establishments began to integrate practical MT into their curricula." (2020: 5). However, if we take a closer look into a paper written by Stephen Doherty about translation technologies and their impact on translation industry, we can find that "...machine translation (MT) had started to develop in the 1930s in the form of mechanical multilingual dictionaries. However, it was not until the 1950s that MT enjoyed a more public showcase as a limited, controlled, but arguably automated translation process." (2016: 6) In simple terms, machine translation is a group of processes, a system per se, that uses complex algorithms to translate a certain amount of text from one source language to another target language while retaining the original meaning and using the provided context accurately. On the official Microsoft website, a more detailed explanation of machine translation can be found, as well as a claim that "Although the concepts behind machine translation technology and the interfaces to use it are relatively simple, the science and technologies behind it are extremely complex and bring together several leading-edge technologies, in particular, deep learning (artificial intelligence), big data, linguistics, cloud computing, and web APIs." (*Machine Translation*) Nevertheless, we must not forget that even though all of this sounds incredibly sophisticated, machine translation is prone to making mistakes, which makes the presence of a human translator necessary. Naturally, some of the more complex fields, such as medicine or law, are especially risky and difficult for machine translation services to translate appropriately.

Of course, nowadays, we live in a world with rapid technological advancements where some other processes and systems help human translators achieve the greatest possible quality of translation output. Some of these are translation quality assurance (TQA)[1] and machine translation quality estimation (MTQE)[2] which will be of importance later.

The ongoing debate in the world of translation is whether a certain translation is good or bad. To answer this question, Doherty proposes a solution:

[1] "Translation quality assurance (TQA) is a set of processes that ensures a translated piece of text is accurate, consistent, and reads well for the target audience. Basically, a quality check for translations." (*What is translation quality assurance?)*

[2] "Machine translation quality estimation (MTQE) is an automated method of predicting the quality of results produced by machine translation tools. It replaces traditional evaluation methods to predict quality without manual evaluation from language experts." (*What Is MT Quality Estimation (MTQE) & Why Is It Important?)*

> "Throughout the long-standing debate on what is a good (or bad) translation, I propose that a dichotomy between accuracy and fluency is apparent across translation theory, translation technology, and in the translation industry in one guise or another, where accuracy typically denotes the extent to which the meaning of the source text is rendered in its translation, and fluency denotes the naturalness of the translated text in terms of the norms of that language. The primary goal of assessing translation quality is ensuring that a specified level of quality is reached, maintained, and delivered as part of the translation product." (2016: 12)

The last thing that is also important to understand is that the terms 'machine translation' and 'automated translation' are not interchangeable. *Poeditor* blog illustrates the difference between the two by saying the following:

> "Machine translation is a broader and more common term used to describe the process of translating text or speech using computer algorithms and technology. It encompasses various methods and technologies for automated translation, and the systems can range from simple online translation tools like Google Translate to more complex and specialised translation software used in professional translation services." (*Machine translation: The future of language*)

In contrast to that, they also compare it with the term 'automated translation' by pointing out the main difference:

> "Automated translation is a term that specifically emphasises the automation aspect of the translation process. It highlights the use of technology to perform translations without significant manual intervention. Automated translation can include both rule-based and statistical methods as well as neural machine translation. The key is that it emphasises the automatic nature of the process." (*Machine translation: The future of language*)

To sum it up, machine translation as a term can be used to encompass the meaning of automated translation, but it does not work the other way around. People generally talk about machine translation when they want to say that they rely on technology and its merits while translating texts.

3.1. Computer-Assisted Translation (CAT) Tools

Computer-assisted translation tools, or CAT tools for short, are another great addition to the world of translation. These tools earned their importance as a necessity in today's world of translation and localization by significantly reducing the effort and time one needs to put into a translation process all the while maintaining the same or even greater output quality. Even though every CAT software differs from the next one, they all operate very similarly. With a quick visit to the *Trados'* website, specifically an article in which they talk about and explain CAT tools, one can find the following description:

> "CAT tools are designed to automate translation-related tasks to make translators faster and more efficient. Translators upload the source document to the application and the interface will split it into segments (phrases, sentences or paragraphs) for them to translate. These source segments and their corresponding translations are then stored in a database known as a translation memory and can be recalled later during the translation process. Over time your translation memories will become bigger and will offer more translation matches, significantly improving your efficiency and increasing consistency." (*What is a computer-assisted translation (CAT) tool?*)

Looking back, we could ask a question such as why and how CAT tools were invented? Doherty explains this in his paper where he states that in the 1990s, software companies and other technology-related industries recognised the importance of translating their products to succeed in international markets. These companies started to look for solutions to increase translation productivity and ensure linguistic consistency across various languages. Since computers, computing power, and internet access became more accessible and affordable over time, this need led to the introduction of computer-assisted translation (CAT) tools, which signified a major technological advancement in the modern translation industry (2020: 4). Doherty then explains CAT behind the scenes by writing: "The core of CAT tools is a translation memory (TM), a software program that stores a translator's translated text alongside its original source text, so that these pairs can later be reused in full or in part when the translator is tasked with translating texts of a similar linguistic composition." (2020: 4) In his paper, Stein compared CAT tools to MT, pointing out a significant difference: "An important distinction exists between MT and computer aided translation (CAT). While the (today not that often announced) goal of

MT is a so-called FAHQT (fully automatic high quality translation), in CAT, tools and methods that assist human translators in the translation process are researched and developed." (2018: 8)

3.2. NMT vs. Adaptive Machine Translation

One important thing is to explain the difference between NMT and adaptive machine translation, since they both fall under the machine translation category, yet they are very different from one another. With both of these translation systems being the latest and considered the best in terms of translation quality, it is interesting to observe why adaptive machine translation has not had as much of an impact as NMT did in the years after its release. With previously mentioned memoQ AGT being in the spotlight as the new representative of adaptive machine translation that learns from context, and not from post-editing as explained earlier, we will try to answer the question at hand. But before that, difference in the way these two systems work will be explained.

On their blog, ModernMT described the typical machine translation system development process with the following image:

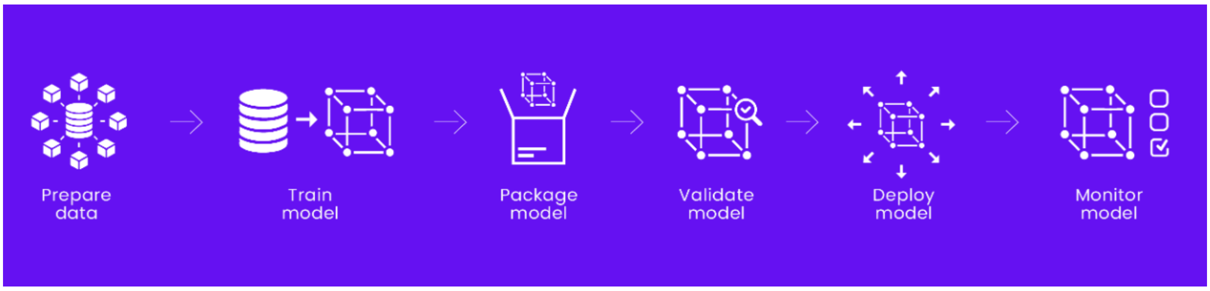

Figure 2: Typical development and production process for developing a static engine

These types of engines do not change the way they would translate a certain sentence, phrase, or term until the next time the engine is updated. This is why we call them static. They are designed for widespread public use, and the cost to build such an engine is very great, as well as the process being extremely complex requiring large amounts of data, which ultimately means it cannot be done on a whim at any given time. Since they require large quantities of data in order to improve, they cannot evolve as quickly. The typical amount of time after an update is done on a static machine translation engine is usually one year. Even though, the use of such engines is so prevalent that we owe 99% of translating being done on Earth to computers, or more precisely, static machine translation engines. (*Understanding Adaptive Machine Translation*)

On the other hand, we cannot simply rely on a generic machine translation engine to be suitable for all purposes and for all enterprises. In order to satisfy the needs of different customers, machine translation models need to be optimised and customised with relevant terminology so they can produce a correct output. This process usually means that the enterprises' translation memory has to be collected and incorporated into a generic machine translation model in a process similar to the generic training process of a machine translation engine. However, this usually results in a limited, roughly optimised model because, if we compare the amount of data used for the optimization to the amount of data used for basic training, the size of the optimization data is minuscule. Additionally, there is little to no value in training such an engine with very limited data because it would have no significant performance differences compared to the baseline model. These problems coupled with the need for building different individual engines for each individual purpose (e.g. marketing, legal, customer support, etc.) that in return need to be optimised separately, make production of machine translation engines and their maintenance a great burden, with the risk of errors and misalignment of data being very high. (*Understanding Adaptive Machine Translation*)

In order to bridge this gap, we can use adaptive machine translation. This approach to translation is perfect and best suited for customers and translators who will want to customise the way the generic model works in order to meet the criteria of use for their business. A simplified way of ongoing processes inside an adaptive machine translation system can be observed on the image below:

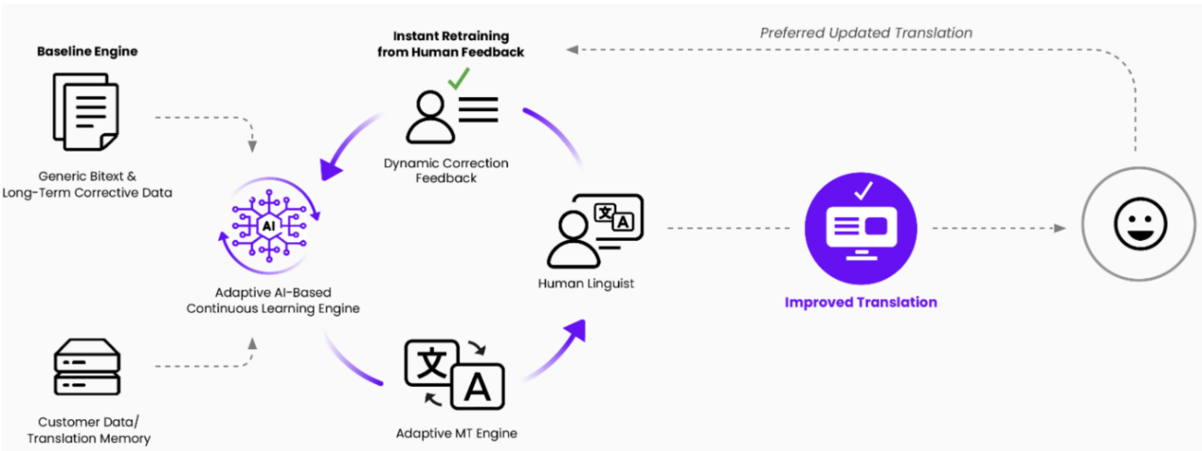

Figure 3: Dynamic and continuously improving adaptive MT and virtuous data network ecosystem

This on-the-go adaptation coupled with constant user feedback is what makes a machine translation engine adaptive. Together with the use of AI, the process is enhanced which results in the productivity of translators being increased. This facilitates the difference between static machine translation engines and adaptive machine translation engines that are designed to excel in tasks and expectations static machine translation engines cannot meet.

A more detailed insight into the specifics of NMT and how it works will be provided in the following chapter, and this segment will be concluded with the following statement: Technology is helpful even when technology itself is the cause of the problem in the first place. The main job and duty of a translator is slowly becoming the accurate and appropriate use of the technologies one has at one's disposal while acting as a critic, judge, and proof-reader of sorts. It has become almost impossible to rely on one’s own skills in the translation industry without the use of technology.

4. Neural machine translation (NMT)

NMT is the main engine behind a lot of different translation services such as Google MT (Google Translate), which was employing the SMT method up until 2016 when the switch to NMT happened. By using neural networks[3] that consist of nodes that imitate a series of neurons in the human brain, it works in such a way that it predicts what is the next most likely word, or a sequence of words, in order to form a sentence. A simplified version of this correlation can be observed in the following image:

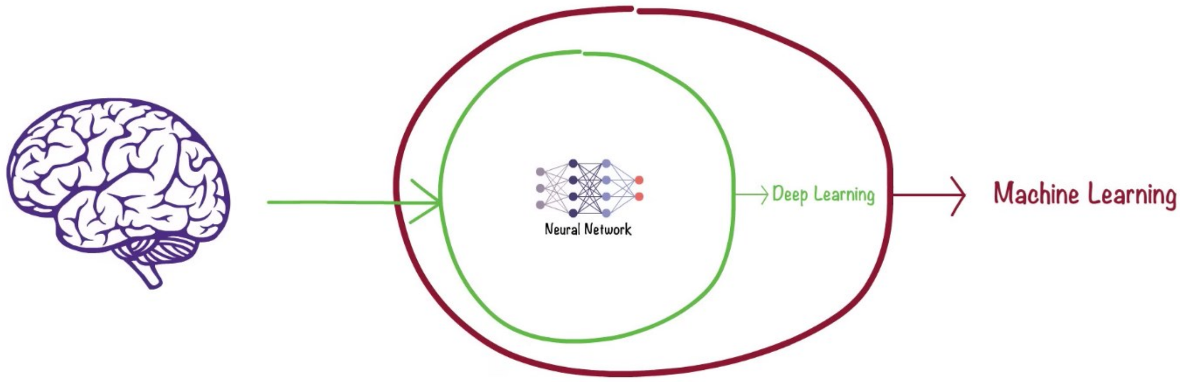

Figure 4: Neural Network In 5 Minutes | What Is A Neural Network | How Neural Networks Work | Simplilearn

Or, to be a bit more precise, each word from the input sentence in the source language is encoded as a number. After the machine provides us with an encoded output that it predicted, each word is decoded back into the target language. (*What is Neural Machine Translation & How does it work?*) To simplify the whole operating process of a neural network, Koehn (2020: 67) explained it as a "...machine learning technique that takes a number of inputs and predicts outputs." Continuing on that, Alsan, in the previously mentioned article for *Weglot,* explains that before it can be used for translation, NMT software must be fed data for training purposes. This data comes in the form of various example translations of a specific text. The software then undergoes training so it can produce an output of the highest quality when given a certain context. (*A Quick Overview of Neural Machine Translation*)

[3] Term used in machine learning; a model based on the structure of a human or animal brain (they take in data which they use for training and recognizing patterns to predict an output)

Because it is the latest and most successful form of machine translation, one can expect that NMT has key advantages over conventional machine translation alternatives. Some of these advantages are higher translation accuracy, reduced reliance on human intervention, and faster completion of translation. Higher accuracy is a direct improvement when compared to previous machine translations since older alternatives lacked the sophistication necessary to accurately translate more complex texts and language pairs. That problem is now officially outdated since NMTs can be trained in order to constantly improve their translation accuracy. As an example of just that, Alsan states that Google, for example, "...found that its Google Neural Machine Translation (GNMT) system reduced translation errors by about 60% compared to its phrase-based production system" (*A Quick Overview of Neural Machine Translation*) after switching to NMT. The second key advantage is pretty evident because a conclusion can be drawn from the previous quote that the output produced by NMT is less time-consuming in the means of post-editing. As for the last key advantage, a perfect explanation on what faster translation actually means can also be found in the same article: "Facebook uses neural machine translation to translate text in posts and comments... While the business previously needed almost 24 hours to train its neural machine translation models, it was able to cut this timeframe to just 32 minutes!" (*A Quick Overview of Neural Machine Translation*)

4.1. Neural Networks

Neural networks, a key component of every NMT software, are the main reason NMT works as well as it does. On *TranslateFXs'* blog, Sam Yip explains that neural networks allow complexity by having a large number of possible parameters, as well as biases and weights, which provide flexibility to the nodes in order to systematise complex data and train complex models. (*What is Neural Machine Translation & How does it work?*) Further detailed information can also be found explaining what exactly neural networks are and how do they work:

> "One way to understand neural networks is to think of the input as a signal with 'information' in it. Some of this information is relevant for the task at hand (e.g. predicting the output). Think of the output as also a signal with a certain amount of 'information'. The neural network attempts to 'repeatedly refine' and compress the input signal's information to match the desired output signal. Think of each hidden layer of the network as stripping out unimportant parts of the input information,

and keeping and/or transforming the output information along the way through the network." (*What is Neural Machine Translation & How does it work?)*

One specific type of neural networks that is also used in LLMs are transformer models. These models are excellent at understanding context, which is crucial for processing human language. By employing a mathematical approach called self-attention, transformer models can identify the nuanced relationships between elements in a sequence. This capability allows them to understand context more efficiently than other machine learning methods. For example, they can recognise how the end of a sentence connects to its beginning and how sentences within a paragraph are related. This, in return, makes an LLM capable of interpreting human language even if that language is of poor quality. (*What is a large language model (LLM)?*)

Based on that knowledge, it is obvious that NMT can be used with any language pair, regardless of the population that that speaks the respective languages since neural networks can be trained with data over time. In addition to that, NMT can also be fine-tuned to produce an output of a particular style, for example scientific, financial, medical, formal, casual and so on. This makes NMT technology the current state of the art technology to be used in the world of translation since it provides the highest quality of translation by far. However, it is indisputable that machine translation still has a long way to go before it can consistently produce high quality output that can match human speech. The question that arises is what can be done to improve such a system? The answer to that question might come on the form of adaptive machine translation powered by AI, namely memoQ AGT. After exploring its capabilities in further detail, this paper will draw a conclusion on performances and accuracy of both memoQ and three other MT engines to determine which one is considered to be 'the best' for the provided context.

5. Machine translation engines

5.1. ModernMT

The first on the list of four machine translation engines is ModernMT, an adaptive neural machine translation service. By browsing through ModernMT's blog, we can find a concise definition of the software as follows: "ModernMT is an MT system that is designed to adapt to the unique needs and focus of an individual translator in essentially the same way that TM does. In many ways, it is a next-generation TM technology that has predictive capabilities." (*MT and the Translator*). Furthermore, ModernMT is an "Instance-Based Adaptive MT" which means it can instantly adapt and tune itself to the subject domain without going through a customization phase, and since it is constantly learning it does not need corrective feedback in order to update and improve. (*A Closer Look at ModernMT*) According to Achim Roupp, who led an independent research in which he compared the outputs of ModernMT against generic MTs and GenAI, it was concluded that ModernMT outperformed all of them in most categories, and proved to have the best output in general (*ModernMT Outperforms Generic MTs and GenAI*). The reason why it is called an 'adaptive' engine is because it adjusts itself on the go as you translate. This was illustrated on ModernMT's blog with the following image:

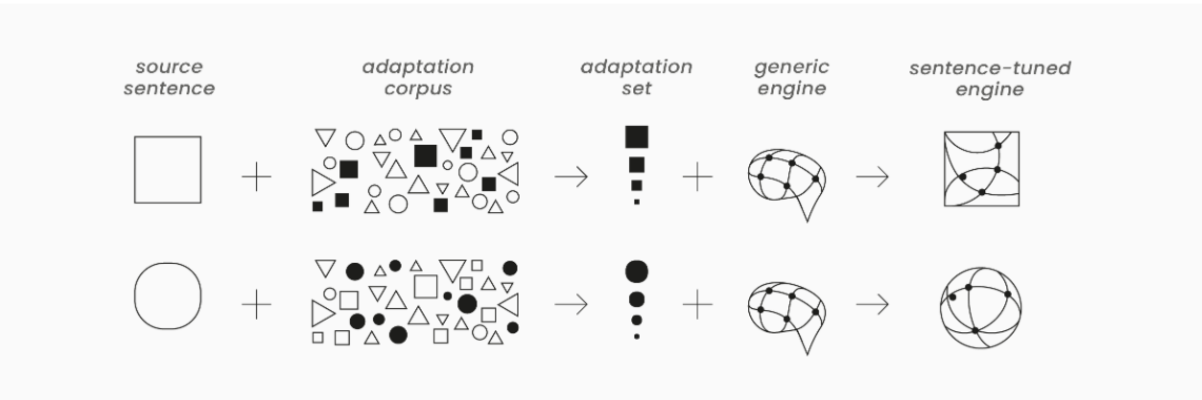

Figure 5: Adaptive MT Adaptation: The ModernMT Scenario

It is designed to leverage the strengths of adaptation to make the work of a professional translator as easy and quick as possible. The main takeaway is that, as opposed to NMT, ModernMT does not need training since it adapts itself on the spot thanks to the translator feedback.

5.2. Google Translate

Arguably one of the most popular translation services in the world, which supports well over 200 languages, is Google Translate. As mentioned earlier, Google Translate has been an NMT engine since 2016, but it certainly existed before that. It was originally released in 2006 as an SMT engine that gathered linguistic data from documents and transcripts provided by the European Parliament as well as the United Nations. If we were to look into the translation process, we could notice that it did not translate directly from the source to the target language. First, it translated the text into English and then used the English version as a source, which was then subsequently translated into the target language. The translation accuracy of Google Translate varies greatly depending on the language pair, but has been improving over the years. The engine is, as one might expect, the most accurate when translating well-structured texts that are put together using not so complex sentences, and are written in a formal tone. Another aspect that significantly raises the quality of translation is whether the topic of the text is well researched. To be more precise, this means that Google Translate has to have access to a large amount of training data related to the topic for the translation quality to increase. Given these circumstances, Google Translate can produce an output similar to that of a human translator.

The importance of Google Translate in everyday life of humans, especially students, is undeniable. A research by Fitrotul Maulidiyah titled *To use or not to use Google Translate* states that "...almost all participants (90%) used GT. Elaboratively, there are 50% of the students who stated that they 'often' use GT, 30% of the students stated they 'sometimes' use GT, 10% of students 'always' use GT, and 10% of students 'seldom' use GT." (2018: 3-4). The researcher also found that the main reasons why students use Google Translate is not only for translation purposes, but writing and vocabulary learning as well. The feedback obtained from the students that participated in this research also confirmed the notion that Google Translate is not perfect because of the fact it sometimes provides an inappropriate translation. Another study carried out by Alla Olkhovska and Iryna Frolova, who examined 48 undergraduate students of translation studies, and the impact MT engines have on their performance, came to the conclusion that "...MT engines for personal use (such as Microsoft Translator) can considerably influence translation students' performance in a negative way, so to improve the results students need to acquire some knowledge about MT and post-editing as well as master practical skills." (2020: 54). Nevertheless, Google Translate and similar engines are very useful tools that can help both students and professional translators alike when used carefully.

5.3. Microsoft Translator

Developed by the Microsoft company, Microsoft Translator is a machine translation service that is a part of their Cognitive Services. Similar to Google Translate, it offers its users translation services while utilizing machine learning and NMT principles explained earlier. This service first became available around the year 2000, and has worked as an SMT engine like the aforementioned Google Translate. In 2016, Microsoft Translator also switched to NMT. The service supports roughly the same number of languages as Google Translate as well as offline translation for certain languages. Among other features worth the mention, it offers speech translation, conversation mode (real-time translation in two languages in case of a language barrier) and camera translation (translating text from pictures or while camera is pointed at a text). All of these features are also available with Google Translate. (*Microsoft Translator vs Google Translate: Translation face-off*)

5.4. MemoQ AGT

By utilizing NMT and AI technology, memoQ AGT presents itself as the most contemporary approach to machine translation currently available in the industry. Generative AI is still a novelty in today's world, and memoQ is taking full advantage of it by combining it with the power of NMT, and additional resources such as TMs, term bases and LiveDocs. A more in-depth discussion about generative AI can be found on memoQ's YouTube channel in a video titled "Large Language Models (LLMs) in Localization - memoQ talks #45", where the host Mark Shriner talks with Kincaid Day and Adam Vance. To sum it up, they defined generative AI as a prompt and response based model that generates content on our behalf. On the other hand, it would be selling it short to say that generative AI is suitable only for content generation. For example, generative AI can take any sort of action we instruct it to, which is a feature of the general intelligence of LLMs absent in machine translation engines (00:18:30). Furthermore, it is impossible to discuss generative AI without mentioning LLM's. If we were to look into further detail on the operation process of memoQ AGT, we could start at the official memoQ website which provides the following description: "memoQ Adaptive Generative Translation is an innovative, AI-driven translation automation technology. It takes translation automation to the next level by using a Large Language Model (LLM) to create translations." (*memoQ AGT; Revamp Your Workflow with the memoQ Ecosystem*). The image from below illustrates this process in a simple yet clear way:

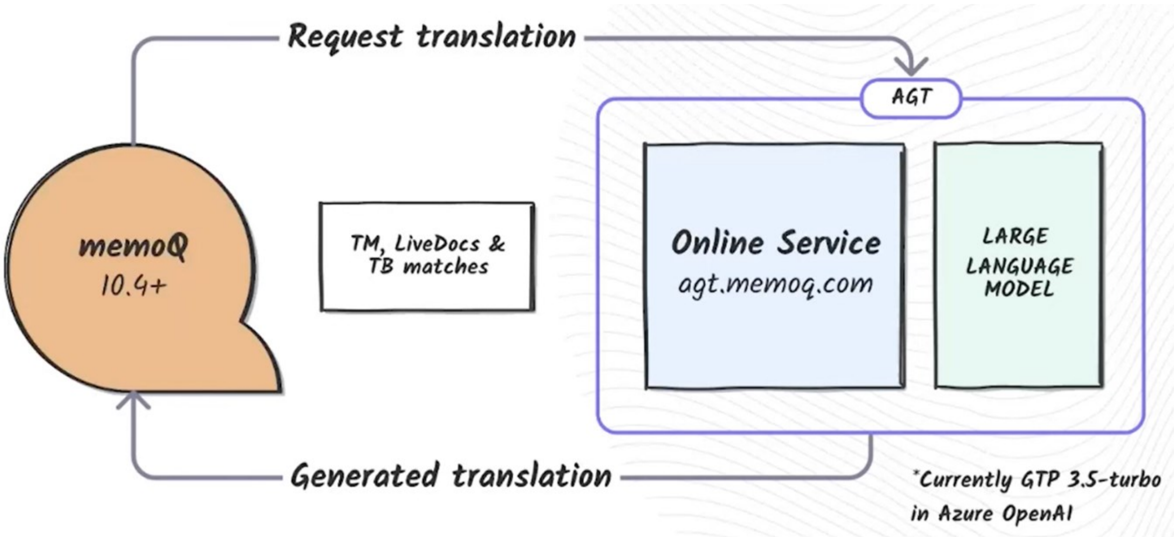

Figure 6: How does AGT work?[4]

If we wanted to put the previous image in words, we could say that the first step implies memoQ sending the text that needs to be translated to AGT. After that, memoQ provides AGT with additional information in the form of translation memory hits and existing term database terms. In the next step, AGT uses this information to put together a prompt for the LLM and sends it to Azure Open AI. Azure Open AI then answers back with a suggested translation as a response to the prompt from AGT. (*memoQ AGT provides AI translation*)

So far, we have discussed machine translation, specifically NMT technology, but the terms that constantly grab the attention are 'AI-driven' and 'Large Language Model' In order to further clarify things, we will explain those terms first. In order to tackle both terms at the same time, LLM will be defined first. To put it simply, on their website *Cloudflare* described an LLM in the following terms: "A large language model (LLM) is a type of artificial intelligence (AI) program that can recognise and generate text, among other tasks. LLMs are trained on huge sets of data — hence the name "large." LLMs are built on machine learning: specifically, a type of neural network called a transformer model." (*What is a large language model (LLM)?*) The most famous example of an LLM could be the ever-present Chat GPT, since it is basically nothing more than an LLM trained with data available up to September 2021. Alongside Chat GPT, many other AI services are currently available thanks to the collaboration between Microsoft Azure and Open AI called Azure Open AI – "...a cloud-based platform that enables developers and data scientists to build and deploy AI models quickly and easily." (*Azure OpenAI Resources*) As it was mentioned earlier, LLMs are not designed specifically for translating; they can perform

[4] Data present on the image is a bit outdated since the current version of memoQ is 11.1 and the LLM is GPT4 and GPT 4o

various tasks, and translation just happens to be one of them. *Medium* wrote a detailed report on Azure Open AI, concluding with the following statement: “Azure Open AI is a powerful platform that provides access to advanced AI models and tools. It can be used to perform a wide range of tasks related to data analysis and processing, including text summarization, question-answering, code generation, and media analysis.” (*Azure OpenAI: A Beginner’s Guide*) They also provided the reader with an image illustrating the way it works:

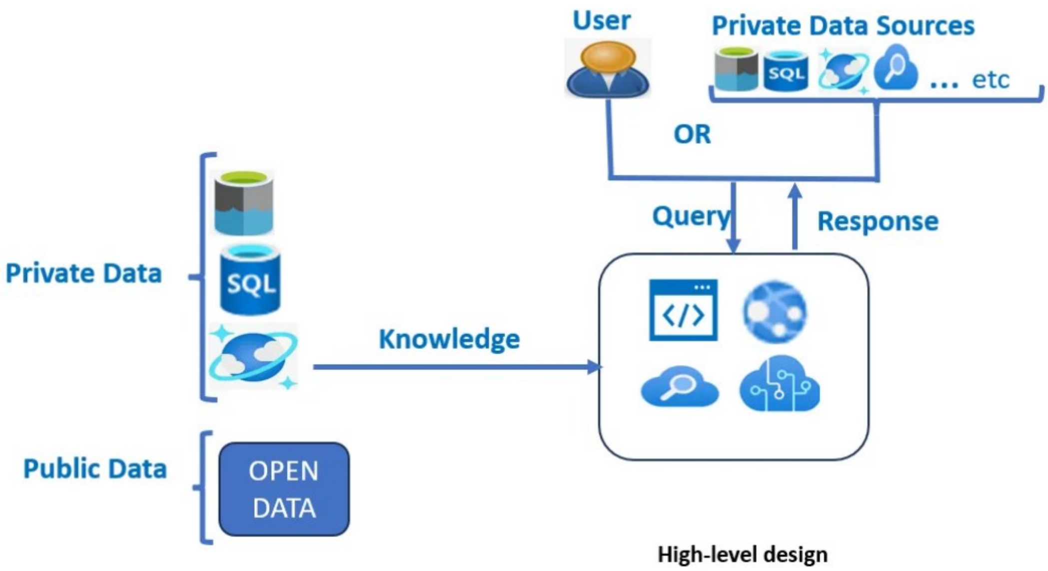

Figure 7: What is Azure Open AI

This is all thanks to the advancements in the field of AI. The way LLMs work is connected to neural networks, which were mentioned and explained in some of the earlier chapters. On their official website, IBM has also done an extensive report on how LLMs work, but the main idea remains that “LLMs operate by leveraging deep learning techniques and vast amounts of textual data. These models are typically based on a transformer architecture, like the generative pre-trained transformer, which excels at handling sequential data like text input. LLMs consist of multiple layers of neural networks, each with parameters that can be fine-tuned during training...” (*What are large language models (LLMs)?*)

However, even though the translation of memoQ AGT is similar to that of Chat GPT by utilizing the power of LLMs, and to that of Microsoft Translator because memoQ is running it as its NMT of choice, the difference between memoQ AGT translation output and the other two translation outputs is memoQ AGT’s instant domain adaptation. This allows memoQ AGT to bypass the long and expensive process of training, which cannot be said for other NMT engines and regular domain adaptation. Agnes Varga, CTO at memoQ, explains this in an online interview on the memoQ YouTube channel by saying that MemoQ AGT learns from context while utilizing resources from the provided translation memories, LiveDocs, and term bases (in case the user was already using memoQ before), which are used to train the LLM to achieve an output of greater accuracy. When using AI for translation, be it either NMT or LLM, it has to be provided with extensive and accurate high-quality data made specifically for the purpose of translation. That way, one can ensure that the AI translation system works smoothly and also decreases the need for post-editing. (*A Closer Look at memoQ AGT*) Currently, memoQ AGT is using the same LLM that Chat GPT is running on, and that is the Azure Open AI 4 and 4o version. This is also corroborated by the founder of memoQ, Gábor Ugray, who stated that "With memoQ AGT, translations are generated by a large language model (LLM), similar to the model powering the popular ChatGPT service. memoQ AGT provides instant domain adaptation to generate translations tailored to the customer's existing language resources, and achieves this without the need to retrain or fine-tune the model itself." (*memoQ to Release Generative AI Translation Automation Tool)*

The real magic of AGT's translation happens when the LLM is provided with context for the example sentence it was instructed to translate. This context are hits from translation memory, term base, and LiveDocs[5], which help the LLM adapt the translation in order to provide the translator with a much more accurate output. This means that memoQ AGT is dependent on data provided by the user, and it works even without being trained, nor do you need to optimise the AI model behind it. This, in return, means that your data will not be made accessible to third parties or external sources, which is important to great majority of people. (*A Closer Look at memoQ AGT*) One obvious implication is that the LLM works based on a short-term memory concept since it will not remember previously translated sentences because the teaching process is segment oriented (00:05:40). This is corroborated by a statement on their official website where it says that AGT "...provides immediate domain customization, ensuring that the translated content aligns with the client's current language assets such as translation memories, terminology, term bases, and aligned documents." (*A Closer Look at memoQ AGT*)

[5] "LiveDocs is a built-in feature of memoQ, an alternative to Translation Memory. With LiveDocs, you can align multiple files in just a matter of seconds and later use them as reference files. As opposed to TM, LiveDocs preserves whole translated documents, not just segments." (*Want to see how LiveDocs works in memoQ?*)

Regarding the quality of translation output, Agnes Varga states that they have experimented with AGT and compared it to a high quality NMT model and with a trained NMT. The comparison was done on memoQ documentation, and the conclusion was that, as expected, where there are many good machine translation hits, the quality of translation exceeds NMT, and if there were not enough hits or if the hits were of lesser quality, then the NMT was superior. (00:15:30) This conclusion was also endorsed by the experience of post-editors. A further clarification of this subject can be found on their website, where they, among other things, discuss the utilization of fuzzy matches as well:

> "memoQ AGT works best for language service providers handling projects with substantial background resources including TMs, terminology, or parallel texts as AGT is fully capable of optimizing the leverage of these resources. If a translator comes across a fuzzy match in a translation memory, it can be tedious to adjust the pre-translated content to fit the current source text, even if the translation in the match is perfect. memoQ AGT excels in this exact scenario. memoQ AGT can also make effective use of so-called low fuzzies. Normally, a human translator would not address a fuzzy match below 80%. However, AGT can work with matches at rates of 50% or 60%, adapting them to the source text." (*A Closer Look at memoQ AGT*)

In conclusion, the size and quality of the TM are crucial to the quality of translation output when translating with memoQ AGT service. This sentiment is also shared by Ugray in a report that can be found on the memoQ website where he states that "An LLM can incorporate these data instantly and on a segment-by-segment basis. memoQ AGT offers the best of both worlds when it combines the generative power of an LLM with the high-quality linguistic data that users of memoQ translation management system (TMS) already have." (*memoQ to Release Generative AI Translation Automation Tool*). In another conclusion, Patrick Molnar, who wrote for their website, also states that a 'human translator stays in the center'. Essentially, memoQ AGT brings about a truly optimal automation in the world of professional translation. Its efficiency is of great assistance to a human translator and is a system that works best exactly when being used by a human translator since it cannot replace one in any way. At the end of the day, even with its near-perfect output (in the right circumstances), there is a need for a human linguist to give the translation a final 'green light' and approve it as valid.

6. Methodology

The source text is an original administrative report on quality assurance in higher education written in the English language. It has to be noted that the source text was drafted by non-native speakers of English, which may have affected its quality, and, subsequently, the quality of machine-translated output. The size of the source text is 27175 characters.

The source text was uploaded to memoQ as a separate project for the purpose of this thesis. The project was set up to include a translation memory containing 9615 entries from previous similar translation projects, a term base containing 1115 verified terms on quality assurance in higher education, and a LiveDocs corpus with 5 reference files provided by the client. The reference files included the site-visit protocol, study programme proposal in English and Croatian, the decision on the appointment of the Expert Panel and the Report form in Croatian, containing the main headings and subheadings.

The source text was then pretranslated using four MT providers, Google Translate, Microsoft Azure AI Translator plugin, ModernMT and memoQ AGT. The pretranslation was done on the entire document at once, without prioritising the available translation memory.

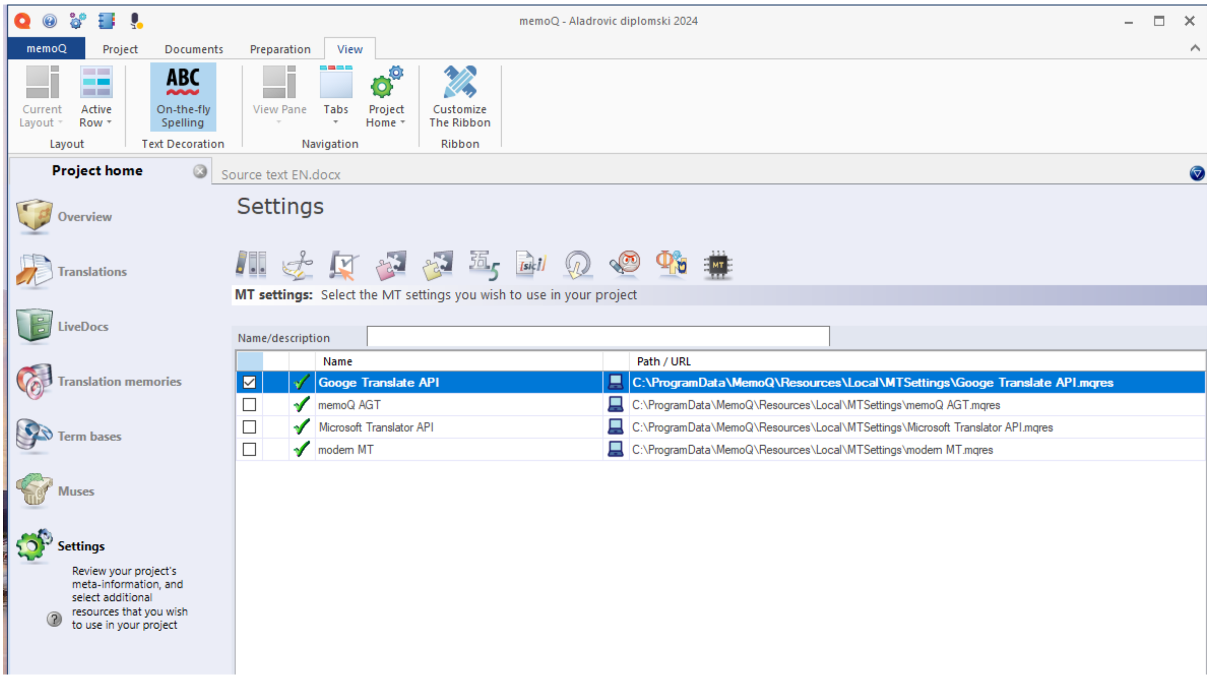

Figure 8: MT plugins and APIs used to pre-translate the source text

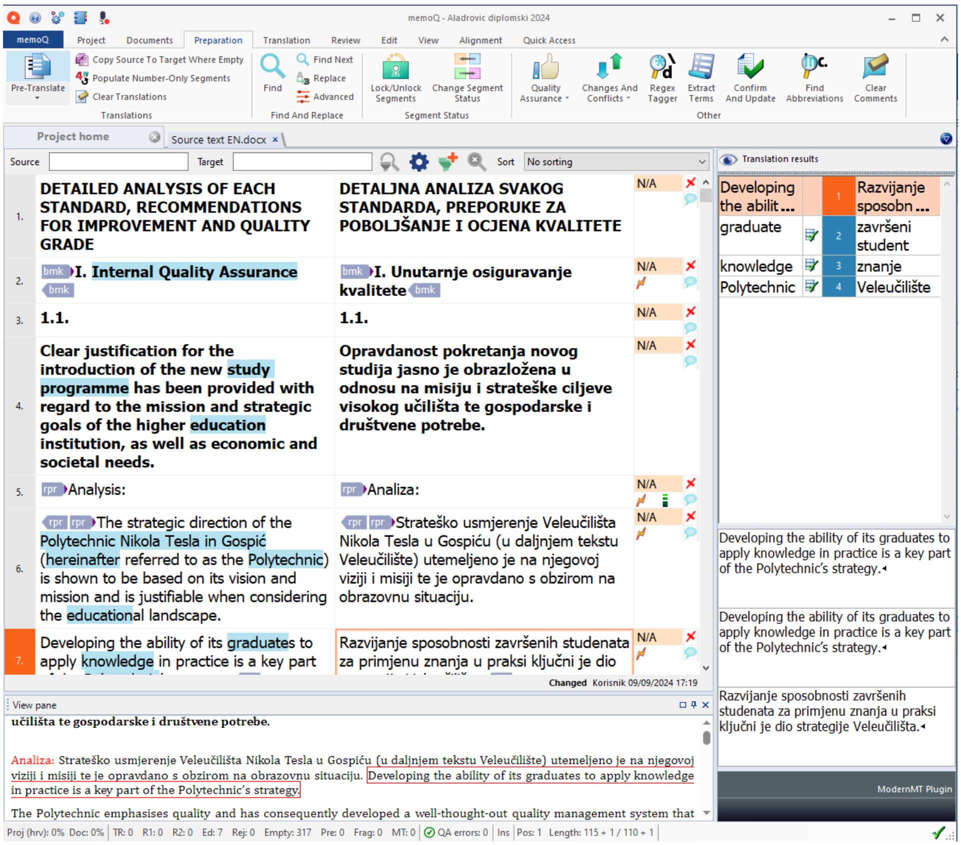

Figure 9: Pre-translation result for ModernMT Plugin

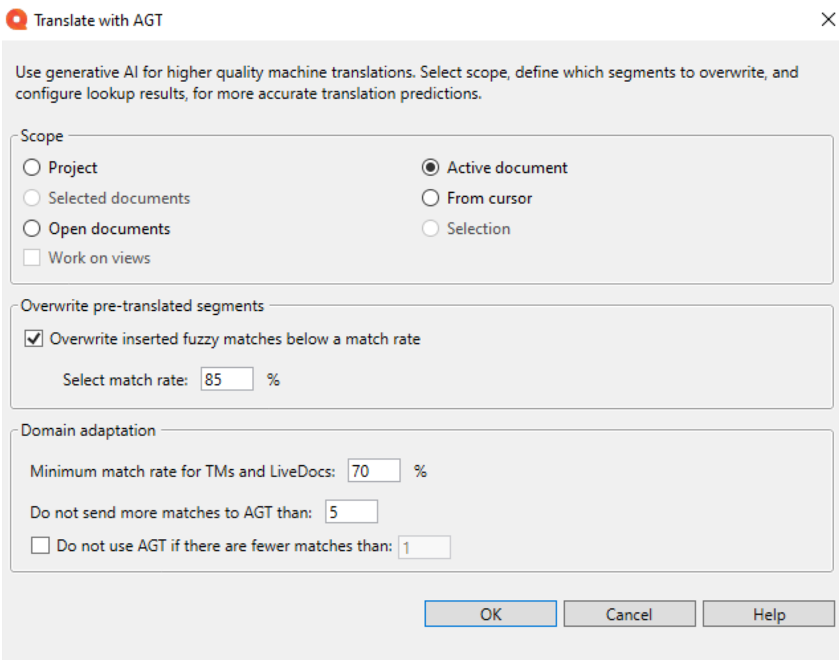

Figure 10: Pre-translation setup for memoQ AGT

The four versions were then compared against each other and the human translation using automatic MT quality assurance metrics provided at the MATEO platform. Furthermore, we performed the qualitative analysis of the four versions against the human translation by a professional translator that we used as the gold standard. Occasionally, however, language-edits of the human translation were suggested were we thought they were necessary. The results and summary of the qualitative analysis of selected examples, as well as MATEO results are presented in Section 7.1 below, as well as in the Conclusion.

6.1. MATEO platform and metrics used

In a paper written on the MATEO platform, Bram Vanroy, Arda Tezcan and Lieve Macken write that MAchine Translation Evaluation Online, or MATEO for short, is an online machine translation evaluation platform. It is essentially a web interface through which users can evaluate machine translation by using different metrics such as BLEU, BLEURT, ChrF, COMET, TER and BERTScore. It is designed to cater to a wide range of users, be it professors, students, researchers, professionals or casual users. The main goal of MATEO is to "...bridge the gap by joining together a diverse set of automatic, reference-based MT evaluation metrics, including both established and cutting-edge methods, into a single, easily accessible web interface" (2023: 499). Among the benefits of using MATEO, some of them are emphasizing the importance of language resources evaluation, improving digital literacy, easy evaluation of machine translation for the purpose of education which helps users critically evaluate MT systems. MATEO is an open-source web interface, and the project was kick-started in 2021 with a Sponsorship grant from the European Association of Machine Translation (2023: 499).

BLUE is the first and most popular of the four metrics that will be used to score the mentioned machine translations. To summarise the significance of this method, Kishore Papineni et al. described BLEU in their paper in the following terms by saying it is a method "...that is quick, inexpensive, and language-independent, that correlates highly with human evaluation, and that has little marginal cost per run." (2002: 311) Furthermore, they write that the human translators "...weigh many aspects of translation, including adequacy, fidelity, and fluency of the translation." (2002: 311) thus making the process, that can take even up to several weeks and months, very expensive which is impossible to sustain in an environment full of daily changes. In order to measure translation performance, two metrics are needed: translation closeness metric as a numerical value and a corpus of high quality human translations. The way BLEU works is that it compares the number of n-gram matches between the reference translation and the candidate translation, and does so for every other candidate translation. It is also important to note that these matches are not position dependent. BLEU simply counts them and decides on the best candidate translation the more matches it has. The final score of the BLEU metric ranges from 0 to 1. (2002: 311-315)

The next metric that will be used is ChrF. In her paper titled “CHRF: character n-gram F-score for automatic MT evaluation”, Maja Popović argues that character n-gram F-score should be used for automatic machine translation output evaluation. Furthermore, she states that “Character n-grams have already been used as a part of more complex metrics, but their individual potential has not been investigated yet.”, emphasising her initial claim. Since it uses the F-score statistic, ChrF, as the name suggests, focuses on characters rather than words. To illustrate the importance of this metric, Popović writes that “...the n-gram based F-scores, especially the linguistically motivated ones based on Part-of-Speech tags and morphemes (Popović, 2011), are shown to correlate very well with human judgments clearly outperforming the widely used metrics such as BLEU and TER.” Similarly to the morpheme F-score, ChrF takes the morpho-syntactic phenomena into account, and it is not dependent on additional tools or sources which means it is language and tokenisation independent. Based on those statements that were confirmed upon experimentation, it was concluded that ChrF is a promising metric since it shows good correlation with human judgment.

The final metric that will be used for machine translation evaluation is COMET. In their paper, Ricardo Rei et al. describe it as “...a neural framework for training multilingual machine translation evaluation models which obtains new state-of-the art levels of correlation with human judgements.” (2020: 2685) They argue that metrics such as BLEU and METEOR are no longer reliable in estimating the quality of MT since “Modern neural approaches to MT result in much higher quality of translation that often deviates from monotonic lexical transfer between languages.” (2020: 2685) They say that the two main challenges MT evaluation faces are accurate segment level correlation with human judgment and adequate differentiation of the highest performing MT systems. As a solution to these problems, they offer COMET, a system that incorporates the source language input into the MT evaluation. Three distinct models that were trained for the purpose of showcasing the advantages of COMET showed promising ability to achieve correlation with human judgments on a segment level.

7. Analysis and discussion

7.1. Qualitative analysis

In this chapter, examples of sentences of each MT output will be compared to one another and one will be chosen as the best solution.[6] The researcher will share their thoughts below every example and shortly explain their decision.

Table 1: Example sentence 1

ST[7]	The external stakeholders met by the Panel provided further evidence of the necessity for a programme of this nature in the region.	TB: External stakeholders – vanjski dionici; the Panel – Stručno povjerenstvo
AGT[8]	Vanjski dionici s kojima se Stručno povjerenstvo susrelo pružili su dodatne dokaze o potrebi za ovom vrstom studijskog programa u regiji.	
MdMT[9]	Vanjski dionici s kojima se Stručno povjerenstvo susrelo pružili su dodatne dokaze o potrebi za ovom vrstom studijskog programa u regiji.	
MT[10]	Vanjski dionici s kojima se sastao panel pružili su dodatne dokaze o potrebi za takvim programom u regiji.	
GT[11]	Vanjski dionici s kojima se susreo Odbor pružili su dodatne dokaze o potrebi za programom ove prirode u regiji.	
POST-EDIT[12]	MdMT / AGT	

[6] At the following link, all of the outputs and the source text document can be found, that were used for this analysis: http://oblak.ffos.hr:8000/d/4867df410056499786c5/?p=%2F&mode=list
[7] Source text
[8] memoQ AGT
[9] ModernMT
[10] Microsoft Translator
[11] Google Translate
[12] Final choice by the post-editor

In this case, AGT and ModernMT provided the best output if we consider the TB hits for the highlighted terms and overall sentence accuracy. Their solution was the best when taking into consideration fluency and structure in the target language. Together with GT they provided the best solution for the term “met by”, unlike Microsoft Translator which provided a suboptimal translation and performed the worst overall.

Table 2: Example sentence 2

ST	The launch of the programme is in alignment with the Polytechnic’s mission and vision and reflects a practical and positive engagement with the local labour market.	
AGT	Pokretanje programa usklađeno je s misijom i vizijom Veleučilišta te odražava praktično i pozitivno sudjelovanje na lokalnom tržištu rada.	
MdMT	Pokretanje programa usklađeno je s misijom i vizijom Veleučilišta te odražava njegovo praktično i pozitivno sudjelovanje na lokalnom tržištu rada.	
MT	Pokretanje programa u skladu je s misijom i vizijom Veleučilišta i odražava praktičan i pozitivan angažman na lokalnom tržištu rada.	
GT	Pokretanje programa u skladu je s misijom i vizijom Veleučilišta te odražava praktičan i pozitivan angažman na lokalnom tržištu rada.	
POST-EDIT	MdMT	

ModernMT has managed to produce the highest quality translation in this case by a narrow margin. AGT and ModernMT had the best choice when translating the term “in alignment” into Croatian. They also had the best choice for translating the term “engagement”, although AGT left out the word “njegovo”, which ModernMT included, that refers to the noun “Veleučilište” and is important semantically. Word choice of GT and MT was inferior in this case, as well as their sentence structure although they did not make any crucial mistakes, and their translation is correct overall but not as good because of the wrong lexical choices.

Table 3: Example sentence 3

ST	It covers crucial areas such as policy development, the establishment of study programmes, teaching methodologies, and ongoing evaluation of all aspects of its activities.	TB: Evaluation – vrednovanje
AGT	Obuhvaća ključna područja kao što su razvoj politika, uspostava studijskih programa, nastavne metode i stalno vrednovanje svih aspekata svojih aktivnosti.	
MdMT	Obuhvaća ključna područja kao što su razvoj politika, pokretanje studijskih programa, nastavne metode i stalno vrednovanje svih aspekata aktivnosti Veleučilišta.	
MT	Obuhvaća ključna područja kao što su razvoj politika, uspostava studijskih programa, metodologije poučavanja i kontinuirana evaluacija svih aspekata njezinih aktivnosti.	
GT	Pokriva ključna područja kao što su razvoj politike, uspostavljanje studijskih programa, metodologija podučavanja i stalna evaluacija svih aspekata njegovih aktivnosti.	
POST-EDIT	MdMT* "stalno vrednovanje" –> "kontinuirano vrednovanje"	

In this example, the results were a bit closer compared to the previous ones since all of the outputs had some issues, as well as correct solutions. Google Translate and Microsoft Translator did not provide a correct translation for the term "evaluation" that exists in the TB. AGT, ModernMT, Microsoft translator gave the plural version of the term "policy" in their output while Google Translate opted for singular, which is not correct. Additionally, all outputs provided a different translation regarding the word "its". The correct solution is not obvious at first from the source sentence example alone, especially because Croatian language is gender sensitive, but when provided with context (previous sentence in the source text) we can see that "its" refers to "Veleučilište" from the previous sentence. ModernMT bypassed this issue by repeating the whole noun which turned out to be the preferred solution. ModernMT was also the only engine that translated the term "establishment" correctly while the other three translated it too literally. Taking everything into consideration, ModernMT has most accurate output thanks to its contextual awareness.

Table 4: Example sentence 4

ST	Students and external stakeholders are involved, and the Polytechnic communicates its QA activities well.	
AGT	U rad su uključeni studenti i vanjski dionici, a Veleučilište uspješno prezentira svoje aktivnosti u području osiguravanja kvalitete.	
MdMT	U rad Odbora uključeni su studenti i vanjski dionici, a Veleučilište uspješno prezentira svoje aktivnosti u području osiguravanja kvalitete.	
MT	Uključeni su studenti i vanjski dionici, a Veleučilište dobro komunicira svoje aktivnosti osiguranja kvalitete.	
GT	Uključeni su studenti i vanjski dionici, a Veleučilište dobro komunicira svoje QA aktivnosti.	
POST-EDIT	MdMT	

AGT and MdMT had the most accurate translation regarding this example, but MdMT gained a slight advantage by translating the start of the sentence with "U rad Odbora..." unlike AGT and its solution where it omitted the word "Odbora" which is important in conveying the meaning of the whole sentence. GT and MT completely omitted that beginning and translated the source sentence too literally which resulted in an inaccurate output that lacks context.

Table 5: Example sentence 5

ST	It is also active in PR activities, providing material to journalists	
AGT	Aktivno je i u odnosima s javnošću te dostavlja materijal novinarima.	
MdMT	Aktivno je i u odnosima s javnošću te dostavlja informacije novinarima	
MT	Također je aktivan u PR aktivnostima, pružajući materijale novinarima	

GT	Aktivan je iu PR aktivnostima, dostavljajući materijale novinarima	
POST-EDIT	MdMT	

In this example, AGT and ModernMT found the right translation for the term "PR activities" which initially made them the preferred choice, but in the second part of the sentence AGT translated "material" too literally whereas ModernMT chose a more suitable translation by outputting "informacije" as a translation for the mentioned term. This gives ModernMT an advantage over AGT and makes it the preferred output. GT and MT on the other hand misgendered "Veleučilište" (which is the noun "aktivno" is referring to) by outputting "aktivan" which is the masculine form of the word thus resulting in inaccurate gender concordance.

Table 6: Example sentence 6

ST	The study programme proposal comprises 18 learning outcomes at the level of the study programme.	TB: (Study programme) proposal – elaborat (studijskog programa)
AGT	Elaborat studijskog programa sadrži 18 ishoda učenja na razini studijskog programa.	
MdMT	Elaborat studijskog programa sadrži 18 ishoda učenja na razini studijskog programa.	
MT	Prijedlog studijskog programa obuhvaća 18 ishoda učenja na razini studijskog programa.	
GT	Prijedlog studijskog programa sadrži 18 ishoda učenja na razini studijskog programa.	
POST-EDIT	AGT / MdMT* sadrži –> sadržava	

In this example there are not many differences in the output. AGT and ModernMT successfully translated the term “study programme proposal” as it can found in the TB, while GT and MT failed to do so. Additionally, MT was the only output that translated the term “comprises” inadequately which makes it the worst output, while translations of AGT and ModernMT are the best overall with a small change regarding the verb “sadrži” which was changed into “sadržava” which is preferred in the Croatian language.

Table 7: Example sentence 7

ST	Learning outcomes are aligned with the level of CroQF.	TB: (Intended) learning outcomes – (predviđeni) ishodi učenja; CroQF – HKO
AGT	Ishodi učenja su usklađeni s razinom CroQF-a.	
MdMT	Ishodi učenja usklađeni su s razinom CroQF-a.	
MT	Ishodi učenja usklađeni su s razinom HKO-a.	
GT	Ishodi učenja usklađeni su s razinom CroQF-a.	
POST-EDIT	MT	

In this example, MT was the only engine that correctly translated the term “CroQF” from the TB. All engines correctly translated the term “learning outcomes” but the only one that had issues with word order is AGT which makes it the worst option in this scenario since this mistake is not negligible given the word order sensitivity of the Croatian language when it comes to the position of the auxiliary verb within the sentence.

Table 8: Example sentence 8

ST	The intended learning outcomes of each course are clearly defined.	TB: Intended learning

		outcomes – predviđeni ishodi učenja
AGT	Jasno su definirani predviđeni ishodi učenja svakog kolegija.	
MdMT	Jasno su definirani predviđeni ishodi učenja svakog kolegija.	
MT	Planirani ishodi učenja svakog kolegija jasno su definirani.	
GT	Predviđeni ishodi učenja svakog predmeta jasno su definirani.	
POST-EDIT	AGT / MdMT	

AGT and ModernMT had the preferred sentence structure in this example as opposed to GT and MT, which made their output the final post-edit choice. On the other hand, GT was the only output that translated the term "course" incorrectly while MT was the only one that incorrectly translated the term from the TB by writing "planirani" instead of "predviđeni".

Table 9: Example sentence 9

ST	The intended course outcomes include the development of both generic and professional competencies.	TB: professional – stručno
AGT	Predviđeni ishodi učenja kolegija uključuju razvoj generičkih i stručnih kompetencija.	
MdMT	Predviđeni ishodi učenja kolegija uključuju razvoj generičkih i stručnih kompetencija.	
MT	Predviđeni ishodi kolegija uključuju razvoj generičkih i profesionalnih kompetencija.	
GT	Predviđeni ishodi predmeta uključuju razvoj generičkih i profesionalnih kompetencija.	
POST-EDIT	AGT / MdMT	

Once again, AGT and MdMT have the identical correct translation of the example sentence. GT translated the term "course" incorrectly as was the case in the previous example, while the only difference that disqualified the output of MT was the exclusion of the word "učenja" before "kolegija". This is important to include because with it the sentence is more accurate semantically since "ishodi učenja" is a common collocation in the Croatian language.

Table 10: Example sentence 10

ST	The HEI accepted the mandatory recommendation of the expert panel and adjusted the study programme proposal accordingly.	TB: HEI – visoko učilište; expert panel – stručno povjerenstvo study programme proposal – elaborat studijskog programa
AGT	Visoko učilište prihvatilo je obveznu preporuku stručnog povjerenstva i sukladno tome prilagodilo elaborat studijskog programa.	
MdMT	Visoko učilište prihvatilo JE obveznu preporuku stručnog povjerenstva i sukladno tome prilagodilo prijedlog studijskog programa.	
MT	Visoko učilište prihvatilo je obveznu preporuku stručnog povjerenstva i u skladu s tim prilagodilo prijedlog studijskog programa.	
GT	Visoko učilište prihvatilo je obveznu preporuku stručnog povjerenstva i sukladno tome prilagodilo prijedlog studijskog programa.	
POST-EDIT	AGT	

MdMT seems to have capitalized "JE" for some unknown reason, otherwise it has the same translation as GT which is mostly correct except for one small difference within the TB hit where it wrote "prijedlog" instead of "elaborat". MT underperformed regarding word choice quality ("u skladu s tim" instead of "sukladno tome"). However, AGT provided the only fully correct translation of the term "study programme proposal" which made its output the preferred one. GT had the same issue as MdMT.

Table 11: Example sentence 11

ST	Examples of these skills are product development, creativity, promotion, innovation, strategic management, leadership.	
AGT	Primjeri tih vještina su razvoj proizvoda, kreativnost, promocija, inovacija, strateško upravljanje, vođenje.	
MdMT	Primjeri tih vještina su razvoj proizvoda, kreativnost, promocija, inovacija, strateško upravljanje, vođenje.	
MT	Primjeri ovih vještina su razvoj proizvoda, kreativnost, promocija, inovacije, strateški menadžment, vodstvo.	
GT	Primjeri ovih vještina su razvoj proizvoda, kreativnost, promocija, inovacija, strateško upravljanje, vodstvo.	
POST-EDIT	Te su vještine primjerice razvoj proizvoda, kreativnost, promocija, inovacija, strateško upravljanje i vođenje.	

AGT and MdMT produced the best output because they opted for the translation "tih" for the word "these" from the source example instead of "ovih" which was the case with GT and MT. AGT and MdMT also translated the term "leadership" correctly unlike GT and MT, while MT was the only output that translated the term "strategic management" incorrectly which made it the worst option. However, the original translator decided to further change the final translation making it more fluent and natural.

Table 12: Example sentence 12

ST	The number of ECTS allocated for the internship is increased from 240 hours to 300 hours and be worth 10 ECTS.	TB: Internship – (stručna) praksa
AGT	Broj ECTS bodova dodijeljenih za praksu povećava se s 240	

	sati na 300 sati i iznosi 10 ECTS bodova.	
MdMT	Broj sati prakse povećava se s 240 na 300 sati i iznosi 10 ECTS bodova.	
MT	Broj ECTS-a dodijeljenih za praksu povećan je s 240 sati na 300 sati i vrijedi 10 ECTS-a.	
GT	Broj ECTS-a koji se dodjeljuje praksi povećava se s 240 sati na 300 sati i vrijedi 10 ECTS-a.	
POST-EDIT	MdMT* Broj sati prakse povećava se s 240 na 300 ~~sati~~ i iznosi 10 ECTS bodova.	

When considering this example sentence it is noticeable that the source text was not put together by native speakers of English language since the meaning of the example sentence is a bit unclear. However, the only translation that has the correct structure which semantically makes the most amount of sense is MdMT. Word structure "povećava se", as opposed to the word structure "povećan je", which is only present in the translation of MT, is correct because this document as a whole talks about actions that will be undertaken and not about those that had already happened. Additionally, word structure "ECTS bodova" (AGT and MdMT) is preferred over "ECTS-a" (GT and MT). Even with a poor quality source example, MdMT managed to provide an excellent translation.

Table 13: Example sentence 13

ST	Requirements and criteria as well as the admissions procedure are clearly defined and transparent.	TB: Requirements – uvjeti; Admissions procedure – postupak upisa
AGT	Uvjeti upisa i kriteriji, kao i postupak upisa, jasno su definirani i transparentni.	
MdMT	Uvjeti upisa, kriteriji upisa i postupak upisa na studij jasno su	

	definirani i transparentni.	
MT	Zahtjevi i kriteriji kao i postupak upisa jasno su definirani i transparentni.	
GT	Zahtjevi i kriteriji kao i postupak upisa jasno su definirani i transparentni.	
POST-EDIT	MdMT* Uvjeti ~~upisa~~, kriteriji ~~upisa~~ i postupak upisa na studij jasno su definirani i transparentni.	

AGT and ModernMT translated the term “requirements” correctly unlike GT and MT. AGT has better sentence structure compared to GT and MT but even so, ModernMT translated it even better. Semantically it makes the most amount of sense and is the clearest meaning-wise within the context of the Croatian language. The original translator thus chose the ModernMT’s translation, but the post-editor decided to left out the unnecessary repetition of the word “upisa”.

Table 14: Example sentence 14

ST	An important point raised is the absence of local entrepreneurs as guest lecturers, even though many are willing and capable of sharing their experiences.	
AGT	Važna točka koja se ističe je odsutnost lokalnih poduzetnika kao gostujućih predavača, iako su mnogi voljni i sposobni podijeliti svoja iskustva.	
MdMT	Stručno povjerenstvo ističe da lokalni poduzetnici nisu uključeni kao gosti predavači, iako su mnogi voljni i sposobni podijeliti svoja iskustva.	
MT	Važna točka koja se postavlja je odsutnost lokalnih poduzetnika kao gostujućih predavača, iako su mnogi voljni i sposobni podijeliti svoja iskustva.	
GT	Važna točka koja se ističe je odsutnost lokalnih poduzetnika kao gostiju predavača, iako su mnogi voljni i sposobni podijeliti svoja iskustva.	

POST-EDIT	MdMT* “uključeni” –> “angažirani”	

MdMT has by far the best solution in this case because it completely changed the sentence structure and word order, so it would make sense when translated into Croatian. The rest suffer from translating the source too literally, and by doing so GT specifically translated the term “guest lecturers” incorrectly so it turned out lecturers had guests which is an incorrect interpretation of the term and an example of incorrect disambiguity. This makes it the worst option, while ModernMT is undoubtedly the best.

Table 15: Example sentence 15

ST	Furthermore, it is noted that some courses lack the integration of entrepreneurial skills.	
AGT	Nadalje, primjećuje se da na nekim kolegijima nedostaje integracija poduzetničkih vještina.	
MdMT	Nadalje, primjećuje se da na nekim kolegijima nedostaje integracija poduzetničkih vještina.	
MT	Nadalje, primjećuje se da nekim tečajevima nedostaje integracija poduzetničkih vještina.	
GT	Nadalje, primjećuje se da nekim tečajevima nedostaje integracija poduzetničkih vještina.	
POST-EDIT	AGT / MdMT* Without “integracija”	

GT and MT continue to have suboptimal translation choices compared to AGT and ModernMT. They seem to fail in grasping the broader context and maintaining consistency since, in this case, they translated the term “course” as “tečaj”, which is incorrect but different from last time when GT offered the term “predmet” as a solution which was also incorrect. Post-editor chose to leave out the word “integracija” when translating the term “integration of entrepreneurial skills” deciding it is unnecessary and the meaning can be conveyed accurately without it.

Table 16: Example sentence 16

ST	Oral exams have been introduced as part of the assessment of student knowledge.	
AGT	Uvedeni su usmeni ispiti kao dio ocjenjivanja znanja studenata.	
MdMT	Uvedeni su usmeni ispiti kao dio provjere znanja studenata.	
MT	Usmeni ispiti uvedeni su u sklopu procjene znanja učenika.	
GT	Uvedeni su usmeni ispiti kao dio provjere znanja studenata.	
POST-EDIT	MdMT / GT	

GT managed to produce a high quality output this time around, as did ModernMT. On the other hand, AGT and MT had issues. AGT opted for "ocjenjivanje" as a solution when translating the term "assessment" which proved to be a suboptimal choice, while MT translated the term as "procjena" which is a different translation but still shares the same issue as AGT of it being suboptimal.

Table 17: Example sentence 17

ST	Mandatory integration of both qualitative and quantitative methods in a module is crucial for students.	
AGT	Obavezna integracija i kvalitativnih i kvantitativnih metoda u modul je ključna za studente.	
MdMT	Obavezna integracija i kvalitativnih i kvantitativnih metoda u modul je ključna za studente.	
MT	Obvezna integracija kvalitativnih i kvantitativnih metoda u modul ključna je za studente.	
GT	Obavezna integracija i kvalitativnih i kvantitativnih metoda u modul je ključna za studente.	
POST-EDIT	MT* Obvezna integracija i kvalitativnih i kvantitativnih metoda u modul ključna je za studente.	

In this example sentence, all four engines performed similarly. We can either choose the output of AGT, MdMT and GT and add an extra "i" before "kvalitativnih", or we can opt for MT which had the correct word order for the auxiliary verb in the term "ključna je" that follows the rules of Croatian language, but omitted the additional "i". Conclusively, it is better to choose MT as the final output since the mistake of putting "je" before the verb "ključno" is a bigger mistake than not adding the conjunction "i" before "kvalitativnih".

Table 18: Example sentence 18

ST	This mandatory approach ensures a well-rounded preparation for students, enhancing their ability to conduct thorough and rigorous research for their final theses.	
AGT	Ovaj obvezni pristup osigurava dobro zaokruženu pripremu za studente, poboljšavajući njihovu sposobnost provođenja temeljitog i rigoroznog istraživanja za njihove završne radove.	
MdMT	Ovaj obvezni pristup osigurava dobro zaokruženu pripremu studenata, jer poboljšava njihovu sposobnost provođenja temeljitog i rigoroznog istraživanja za završne radove.	
MT	Ovaj obvezni pristup osigurava dobro zaokruženu pripremu za studente, poboljšavajući njihovu sposobnost provođenja temeljitog i rigoroznog istraživanja za svoje završne radove.	
GT	Ovaj obvezni pristup osigurava dobro zaokruženu pripremu za studente, poboljšavajući njihovu sposobnost provođenja temeljitog i rigoroznog istraživanja za svoje završne radove.	
POST-EDIT	MdMT	

Once again, the output of ModernMT sounds the most natural in Croatian language compared to the other three. It adjusts the translation precisely for that purpose while the rest, as was already said, translate a bit too literally. This adjustment makes it semantically superior to the translations of the other three engines.

Table 19: Example sentence 19

ST	At the moment of initial accreditation, the Polytechnic is subsidising costs of tuition on postgraduate (doctoral) studies for their two teachers.	
AGT	U trenutku inicijalne akreditacije, Veleučilište subvencionira troškove školarine na poslijediplomskom (doktorskom) studiju za svoja dva nastavnika.	
MdMT	Veleučilište u trenutku inicijalne akreditacije subvencionira troškove školarine na poslijediplomskom (doktorskom) studiju za svoja dva nastavnika.	
MT	U trenutku inicijalne akreditacije Veleučilište subvencionira troškove školarine na poslijediplomskom (doktorskom) studiju za svoja dva nastavnika.	
GT	Veleučilište u trenutku inicijalne akreditacije subvencionira troškove školarine na poslijediplomskom (doktorskom) studiju za svoja dva nastavnika.	
POST-EDIT	MdMT / GT	

Even though the example sentence is written as an inversion, that does not mean the translation in Croatian needs to be an inversion as well. ModernMT and GT recognized this and translated the sentences so it sounds more natural, which would not be the case if it was an inversion as was the case with the outputs of AGT and MT, further proving the notion that they are translating too literally.

Table 20: Example sentence 20

ST	Since the new annex is still not finished and new facility of the library is not finished, and new equipment is not available to see, the Panel recommends finishing all construction works and to equip the rooms as planned.	TB: the Panel – Stručno povjerenstvo
AGT	Budući da novo krilo još uvijek nije dovršeno, a novi prostor	

	knjižnice nije dovršen, a nova oprema nije dostupna za pregled, Stručno povjerenstvo preporučuje dovršenje svih građevinskih radova i opremanje prostora prema planu.	
MdMT	Budući da novo krilo i novi knjižnični prostori još uvijek nisu dovršeni, a nova oprema nije dostupna za pregled, Stručno povjerenstvo preporučuje dovršenje svih građevinskih radova i opremanje prostora prema planu.	
MT	Budući da novi aneks još uvijek nije dovršen, a novi objekt knjižnice nije dovršen, a nova oprema nije dostupna za razgledavanje, Povjerenstvo preporučuje završetak svih građevinskih radova i opremanje prostorija prema planu.	
GT	Budući da novi aneks još uvijek nije završen i novi objekt knjižnice nije završen, a nova oprema nije dostupna za razgledavanje, Vijeće preporučuje završetak svih građevinskih radova i opremanje prostorija prema planu.	
POST-EDIT	MdMT	

It continuously seems that the only engine that has no problem adapting the translation to fit the Croatian language and make it sound the most natural is ModernMT. The rest persistently struggle with more complex sentences or when the source text is of lesser quality (once again the example sentence is obviously not written by native English speakers). In addition to that, MT and GT once again failed to translate the term “the Panel” from the TB as “Stručno povjerenstvo”. Considering the fact MdMT outperformed AGT in the first part of the sentence, it is once again the preferred choice. AGT piled up conjunctions by literally translating the example sentence which made the translation sound unnatural and less fluent compared to MdMT

7.2. Quantitative analysis

In example sentence 1 AGT and ModernMT performed the best if we consider the TB hits, fluency and word choice. Their translation left nothing up to debate and was accepted completely. On the other hand, MT and GT had suboptimal word choice and lacked the context needed to produce a sentence that would be sufficiently fluent. The same was true for example sentence 2 although this time AGT performed slightly worse overall by not including an important possessive pronoun for the meaning of the part of the sentence in question. MT and GT once again had suboptimal translation choices for certain words. However, they gained momentum in example sentences 3 where their output was very close in terms of quality and fluency. Every output had flaws and strong suits in this example, but ModernMT still managed to come out on top with just one minor adjustment by the original translator making its output the most correct overall. ModernMT also performed best in terms of contextual awareness because the rest of the MT engines struggled to translate the word "its" into Croatian. The solution for this issue can be found if the MT engine is capable of learning from the whole text instead of being sentence, or segment, based. Sentence example 4 provides us further insight into the issue of translating too literally. This issue persists in the translations of MT and GT. On the other hand, AGT had problems with it rarely while ModernMT dominated with its performance by not scoring any negative points regarding this issue. This advantage is what made ModernMT's translation the best in the following example as well, even though it was by a narrow margin. This example sentence is also the first time GT and MT had issues with genders since they misgendered a noun. This could have been avoided if they, as was mentioned earlier, had the ability to look into the previous sentences and use the broader context of the whole document to their advantage. Translations for the example sentence 6 were very similar. AGT and ModernMT provided a more accurate translation because their translation followed the TB solution for a given term, and they had a more adequate solution when it came down to a certain word later in the sentence. An upset followed in the example sentence 7 where MT had the most accurate translation. Even though it was a simple, and short sentence, MT was the only engine that correctly translated the term from the TB while others failed to do so. In addition to that, AGT performed the worst by messing up the word order. As it is usually the case regarding the translation of example sentences observed in this analysis, ModernMT and AGT perform similarly with ModernMT having a slight advantage, and MT and GT on the other hand performing slightly worse but comparable to one another. This was best illustrated in the translation of the example sentence 8 where AGT and ModernMT had the preferred sentence structure over MT and GT, providing us with a perfect translation. However, GT performed the worst if we consider the mistake of translating a term from the TB incorrectly which was not the case with other outputs. The struggle to provide the correct translation according to the hits from the TB continues for MT and GT in example 9. AGT and ModernMT showed no mistakes in their translations. Example sentence 10 delivered another upset because this time ModernMT provided a wrong translation for a TB term which it translated correctly just a few sentences before. AGT did not make this mistake and because of that it was chosen as the preferred translation. MT and GT made the same mistake, and additionally, MT struggled with word choice which is a common issue throughout the analysis. In the following example, we noticed that MT and GT, among other, struggle with demonstrative pronouns when translating into Croatian language, while AGT and ModernMT hold their ground by providing an adequate translation. Nevertheless, the original translator further changed the output to reach the target fluency and naturalness of the target language. Example sentence 12 was one of the sentences where it was noticeable that the source text was not put together by a native speaker of English language, and since translating too literally and choosing inadequate solutions were common issues for MT and GT specifically, this was the case once again in this example. They were joined by AGT in their mistakes, but ModernMT managed to put together a decent translation despite the mentioned problem which was ultimately, with a minor adjustment, chosen as the final translation. Example sentence 13 did not provide any important new insights regarding weaknesses of a specific MT output. ModernMT once again had the preferred output which was trimmed down a bit by the translator who decided to leave out the unnecessary repetitions to make the translation more fluent. Example sentence 14 further proves that ModernMT has the best contextual awareness and the ability to not only translate but adapt the translation to fit the fluency and style of the target language. AGT, MT and GT once again fell short in that regard. Additionally, AGT and MT demonstrated a case of incorrect disambiguity in the translation of said example because they conveyed the meaning of a certain term in the wrong way. GT had a similar issue while ModernMT had no issues whatsoever. In example sentence 15, MT and GT once again demonstrate a lack of consistency and contextual awareness. Additionally, GT once again translated a specific term that appeared in one of the previous examples incorrectly, but this time it chose a different incorrect translation. By deciding to leave out a single word which was deemed unnecessary, the translator chose the translations of AGT and ModernMT as the final solution. Example sentence 17 is the first and last one where GT came up with a completely accurate translation together with ModernMT. It correctly translated the highlighted terms and had correct collocations in target language. AGT and MT on the other hand had suboptimal word choices. Example sentence 17 once again proved MT should not be completely disregarded since it produced the best translation by having the correct word order in the target language. This was interesting since a few examples prior to this one ModernMT and GT did not make this mistake, while AGT is consistently having the incorrect word order when it comes to auxiliary verbs. In the next example, ModernMT and its ability to localize the translation shines once again compared to suboptimal sentence structure and word forms of AGT, MT and GT. Example sentence 19 is a perfect illustration how AGT and MT can be wrong in the same way, which is expected since AGT is based on MT. Again, they translated too literally and copied the structure of the source language which oftentimes does not work when translating from English into Croatian. GT surprisingly did not make the same mistake and provided the same output as ModernMT which was correct and accurate. The last example sentence proved the notion that was repeated throughout the analysis, and that is AGT's inability to localize the translation and make it sound natural in the target language, at least when that language is Croatian. ModernMT managed to translate the sentence appropriately even though the source example was once again questionable at best in terms of structure. MT and GT continue to provide translations of lesser quality for specific terms with suboptimal sentence structure in the target language.

When taking all examples into consideration, we can conclude that AGT still has room for improvement before it can be considered the best. ModernMT has an advantage in producing a more fluent and meaningful translation when the target language is Croatian. AGT on the other hand struggles with that even though it provides better solutions for certain terms when compared to GT and MT. However, although rare, there are certain cases where GT or MT provided the preferred output. Nevertheless, they struggle the most when it comes to translating terms that require contextual awareness and have difficulty adapting the translation, so it sounds fluent and natural in the target language.

Next figures will show the results of the MATEO platform analysis regarding the specific metrics chosen (BLUE, ChrF and COMET). All four MT outputs were given a certain score for a specific metric. The numerical value indicates how well they did on each test with the highest score being the best. First two figures illustrate the same score distribution chart, just depicted differently, and the third figure shows individual cases in each translation output that were scored individually.

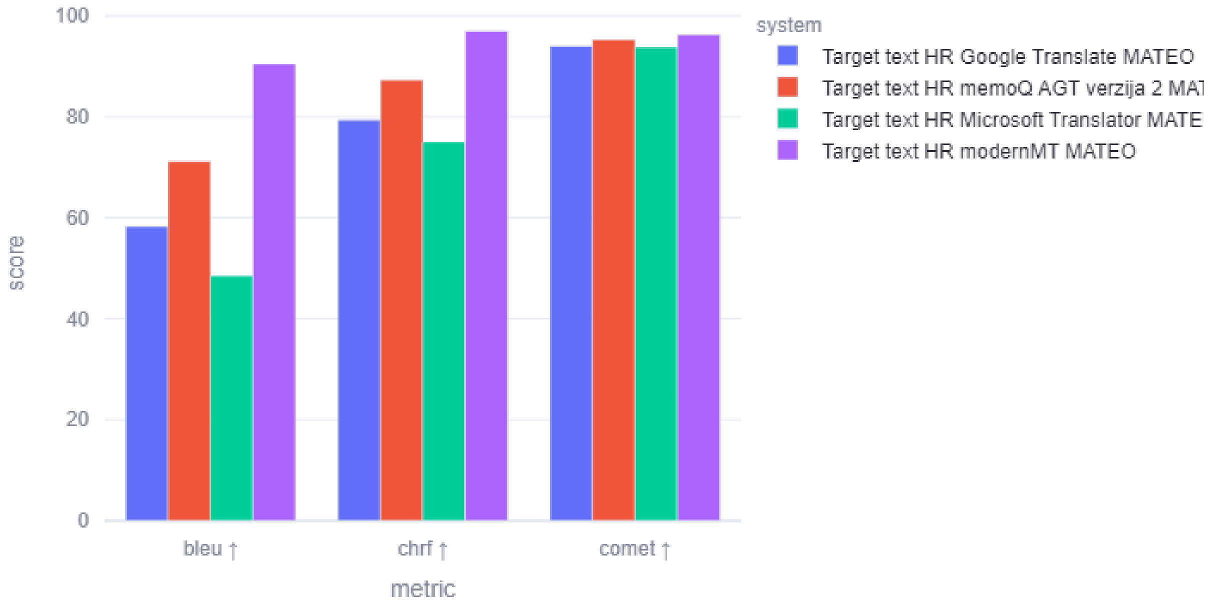

Figure 11: MATEO metrics results 1

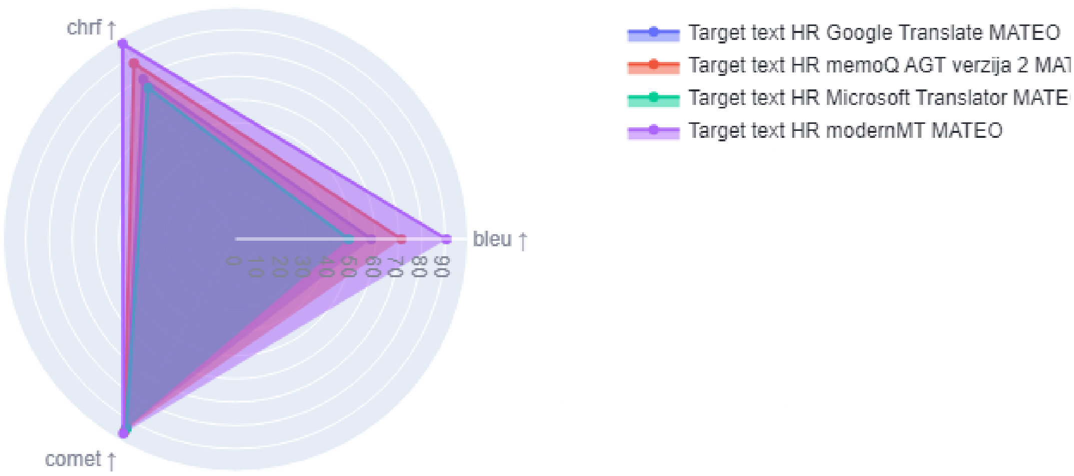

Figure 12: MATEO metrics results 2

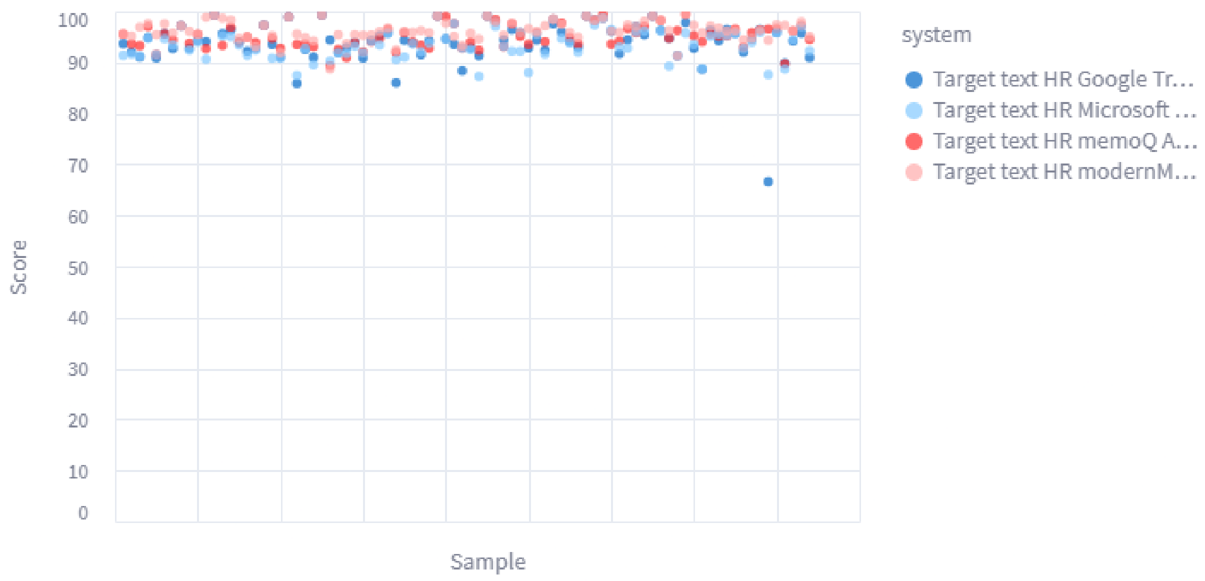

Figure 13: MATEO metrics results 3

As it is visible from the Figures that represent MATEO metrics results, which were listed as ones this analysis will use, the observations in the comments below every table with sentence examples is proven correct. The higher the score of the metric – the better the translation. However, scores between outputs for specific metrics do vary because BLEU and ChrF "...measure the string overlap between the MT outputs and reference translations" while "COMET measures translation quality in a general sense" (*Using MT metrics*). Furthermore, COMET leverages machine learning in its evaluation, and does not rely on superficial text-comparison like more traditional metrics. By taking into consideration elements like fluency, preservation of meaning and adequacy, COMET is most effective in translation cases which require a deeper understanding of quality. BLEU, on the other hand, compares a certain translation to the reference translation and by doing so measures correspondence of phrases between the two, focusing specifically on word matches and is thus very effective in evaluating translations where word order and exact matching of phrases is important (*Using MT metrics*). The third metric, ChrF, "...is a metric for machine translation evaluation that calculates the similarity between a machine translation output and a reference translation using character n-grams, not word n-grams" (*chrF; CHaRacter-level F-score*). By comparing different figures, it is noticeable that the biggest difference in individual scores is within the BLEU metric. ChrF metric has higher scores overall and less of a deviation, while COMET has the highest scores and the smallest deviation between specific translation outputs. It is important to note the improvement of AGT compared to MT across all metrics since AGT is based on MT, but offers various new improvements. These metrics results prove that AGT indeed is an improvement compared to standard NMT engines.

However, ModernMT is conclusively the best MT out of four that were tested with the highest scores in each metric. AGT outperformed Google Translate and Microsoft Translator but failed to reach the heights of ModernMT. In the English-Croatian language combination, AGT still does not match the results of adaptive NMT (ModernMT), but it certainly is an improvement when compared to standard NMT engines (Google Translate and Microsoft Translator). This could have to do with the limited resources it was provided. If we had had a larger TM or had tweaked the AGT settings to allow for more hits, the results may have been different.

8. Conclusion

When compared to other MT engines, AGT performed well, but not well enough in order to unanimously declare it as the best MT engine. Instant domain adaptation puts it in an advantageous position when compared to GT and MT, but there are still shortcomings in terms of fluency, style and contextual awareness. We can also notice that AGT does take into consideration hits from TB, albeit not consistently. This is still better than not consulting with the TB at all, unlike other engines that were a part of this analysis. Overall, ModernMT performed best across the board with AGT coming in close second. GT and MT proved their usefulness, albeit not often enough to be looked upon favourably when put against AGT or ModernMT. AGT certainly has potential, and we could speculate it could perform better in another language combination, but when put in the English-Croatian language combination it underperformed. Currently, AGT offers no significant advantages, at least not when compared to ModernMT, and in the English-Croatian language combination since it oftentimes failed to capture the fluidity of ModernMT and its translations. The glorified instant domain adaptation and short-term memory of the LLM and its segmented, sentence-based teaching process seemed to have hurt its performance in this case, since it often failed to provide the accurate translation which ModernMT excelled at, as was evident from the solutions it provided that were presumably enhanced by its ability to grasp the broader context of the whole document, and not just working with the singled out example sentence. Based on our corpus and research, working in the same setting and the same resources, Adaptive Generative Translation (AGT) shows a clear improvement over standard NMT engines, but still does not seem to match the overall quality of adaptive NMT.

9. Works cited

Alsan, Merve. “What Makes Neural Machine Translation Better Than Other Translation Techniques?” *Weglot,* Weglot, 10 June 2024. Web. 10 August 2024. <https://www.weglot.com/blog/neural-machine-translation#what-makes-neural-machine-translation-better-than-other-translation-techniques>

Buddhabhatti, Alpa. “Azure OpenAI: A Beginner’s Guide.” *Medium*. Medium, 10 February 2023. Web. 10 August 2024. <https://medium.com/@meetalpa/azure-openai-a-beginners-guide-0fca54ee89cf>

Doherty, Stephen. “The impact of translation technologies on the process and product of translation.” *International Journal of Communication 10* (2016), 947–969. *The University of New South Wales, Australia.* Web. 19 August 2024. <https://ijoc.org/index.php/ijoc/article/download/3499/1573>

Harrison, Tovmas. “Statistical Machine Translation.” *Machine Translate*, Machine Translate, 30 October 2023. Web. 13 August 2024. <https://machinetranslate.org/statistical-machine-translation>

Harrison, Tovmas. “chrF; CHaRacter-level F-score.” *Machine Translate,* Machine Translate 30 October 2023. Web. 20 September 2024. <https://machinetranslate.org/chrF#:~:text=CHaRacter%2Dlevel%20F%2Dscore&text=chrF%20was%20introduced%20in%202015,forms%20of%20a%20single%20word.>

“Introducing memoQ Adaptive Generative Translation.” *YouTube,* 3 December 2023. Web. 25 August 2024. <https://www.youtube.com/watch?v=JPdO1Hi__iQ>

Kenny, Dorothy. “Technology in Translator Training.” *The Routledge Handbook of Translation Technology*, *Routledge.* 2020, pp. 498-515. Web. 28 August 2024. <https://www.researchgate.net/publication/339796568_Kenny_Dorothy_2020_'Technology_in_Translator_Training'_In_Minako_O'Hagan_ed_The_Routledge_Handbook_of_Translation_Technology_London_and_New_York_Routledge_498-515>

“Large Language Models.” *IBM*, IBM, n.d. Web. 15 August 2024. <https://www.ibm.com/topics/large-language-models>

"Machine Translation: An Overview." *POEditor*, POEditor, 9 October 2023. Web. 6 August 2024. <https://poeditor.com/blog/machine-translation/>

"Machine Translation." *Microsoft*, Microsoft, n.d. Web. 14 August 2024. <https://www.microsoft.com/en-us/translator/business/machine-translation/>

"Machine Translation Quality Estimation: Why and How to Do It." *POEditor*, POEditor, 8 April 2024. Web. 6 August 2024. <https://poeditor.com/blog/machine-translation-quality-estimation/>

Maulidiyah, Fitrotul. "To Use or Not to Use Google Translate." *JLT – Jurnal Linguistik Terapan, vol. 8, no. 2.* (2018), *Politeknik Negeri Malang.* Web. 23 August 2024. <https://jurnal.polinema.ac.id/index.php/jlt/article/view/178/70>

"memoQ AGT: A Public Plugin for Machine Translation." *memoQ*, memoQ, n.d. Web. 25 August 2024. <https://www.memoq.com/integrations/machine-translation/public-plugin/memoq-agt>

"memoQ AGT with Agnes Varga, CTO at memoQ - memoQ talks #47" *YouTube*, 5 December 2023. Web. 25 August 2024. <https://www.youtube.com/watch?v=pbRrvkAdcRc>

"memoQ to Release Generative AI Translation Automation Tool." *memoQ*, memoQ, 12 October 2023. Web. 25 August 2024. <https://www.memoq.com/pressroom/memoq-to-release-generative-ai-translation-automation-tool-2>

"Microsoft Translator vs Google Translate." *POEditor,* POEditor, 30 October 2023. Web. 22 August 2024. <https://poeditor.com/blog/microsoft-translator-vs-google-translate/>

"ModernMT: A Public Plugin for Machine Translation." *memoQ*, memoQ, n.d. Web. 30 August 2024.<https://www.memoq.com/integrations/machine-translation/public-plugin /modernmt#:~:text=What%20is%20ModernMT%3F,to%20train%20your%20MT%20engine>

Molnar, Patrick. "A Closer Look at memoQ AGT." *memoQ*, memoQ, 14 November 2023. Web. 30 August 2024. <https://blog.memoq.com/a-closer-look-at-memoq-agt>

"MT Quality Estimation: Ensuring Accurate Machine Translation." *translate plus*, Publicis Groupe, n.d., Web. 8 August 2024. <https://www.translateplus.com/blog/mt-quality-estimation/>

Olkhovska, Alla. Frolova, Iryna. “Using Machine Translation Engines In The Classroom: A Survey Of Translation Students’ Performance.” *Advanced Education 7 (15): p.p. 47-55.* (2020), *V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.* Web. 30. August 2024. <https://www.researchgate.net/publication/343120981_USING_MACHINE_TRANSLATION_ENGINES_IN_THE_CLASSROOM_A_SURVEY_OF_TRANSLATION_STUDENTS'_PERFORMANCE>

“OpenAI Resources.” *Microsoft*, Microsoft 21 May 2024. Web. 10 August 2024. <https://microsoft.github.io/PartnerResources/skilling/ai-ml-academy/resources/openai>

Papineni, Kishore. et al. “BLEU: a Method for Automatic Evaluation of Machine Translation.” *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, 2002.* pp. 311-318. Web. 12 August 2024. <https://aclanthology.org/P02-1040/>

Ping, Ke. “Machine Translation.” *Routledge Encyclopedia of Translation Studies, Second edition.* 2009. pp. 162-169. *Routledge.* Web. 6 September 2024 <https://elearning.unipd.it/scienzeumane/pluginfile.php/422897/mod_folder/content/0/Routledge_Encyclopedia_of_Translation_Studies_2nd_ed.pdf>

Popović, Maja. “CHRF: character n-gram F-score for automatic MT evaluation.” *Proceedings of the Tenth Workshop on Statistical Machine Translation.* 2015. pp. 392-395. *Association for Computational Linguistics.* Web. 12 August 2024. <https://aclanthology.org/W15-3049/>

Qiu, Disheng. “ModernMT Outperforms Generic MTs and GenAI.” *MultiLingual*, MultiLingual, n.d. Web. 21 Aug. 2024. <https://multilingual.com/magazine/march-2024/modernmt-outperforms-generic-mts-and-genai/>

Rei, Ricardo. et al. “COMET: A Neural Framework for MT Evaluation.” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).* 2020. pp. 2685-2702. *Association for Computational Linguistics.* Web. 12 August 2024. <https://aclanthology.org/2020.emnlp-main.213/>

Stein, Daniel (2018). “Machine translation: past, present and future.” *Language Science Press: Berlin.* 2013, Web. 28 August 2024. <https://langsci-press.org/catalog/view/106/219/1125-1>

"Translation Quality Assurance: Best Practices for Accurate Localization." *POEditor*, Code Whale Inc., 27 May 2023. Web. 8 August 2024. <https://poeditor.com/blog/translation-quality-assurance/>

"Using MT Metrics" *Phrase,* Phrase, n.d. Web. 20 September 2024. <https://support.phrase.com/hc/en-us/articles/12669609584156-Using-MT-Metrics#:~:text=BLEU%2C%20chrf%20and%20TER%20all,quality%20in%20a%20general%20sense.>

Vanroy, Bram. Tezcan, A., and Macken, L. "MATEO: MAchine Translation Evaluation Online." 2023. *M. Nurminen, J. Brenner, M. Koponen, S. Latomaa, M. Mikhailov, F. Schierl, … H. Moniz (Eds.), Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 499–500). T*ampere, Finland: European Association for Machine Translation (EAMT).* Web. 12 August 2024. <https://lt3.ugent.be/mateo/>

Vashee, Kirti. "MT and the Translator." *ModernMT,* ModernMT, 2 March 2022. Web. 19 September 2024. <https://blog.modernmt.com/mt-and-the-translator/>

Vashee, Kirti. "A Closer Look at ModernMT." *ModernMT,* ModernMT, 27 September 2021. Web. 19 September 2024. <https://blog.modernmt.com/modernmt-a-closer-look-at-an-emerging-enterprise-mt-powerhouse/>

"Want to see how LiveDocs works in memoQ?" *memoQ*, memoQ, n.d. Web. 26 August 2024. <https://www.memoq.com/tools/livedocs>

"What Is a CAT Tool?" *Trados*, Trados, n.d. Web. 15 August 2024. <https://www.trados.com/learning/topic/cat-tool/>

"What Is a Large Language Model (LLM)?" *Cloudflare*, Cloudflare, n.d. Web. 23 August 2024. <https://www.cloudflare.com/learning/ai/what-is-large-language-model/>

Xiao, Haozhe, Y. Zhou, and Y. Lepage. "Example-Based Machine Translation with a Multi-Sentence Construction Transformer Architecture." P*roceedings of the 2023 CLASP Conference on Learning with Small Data (LSD)* pp. 72–80. *Association for Computational Linguistics*, Web. 19 August 2024. <https://aclanthology.org/2023.clasp-1.9/>

Yang, Jingfeng. et al. "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond." *Association for Computing Machinery,* 2024. Web. 30 August 2024. <https://arxiv.org/abs/2304.13712>

Yip, Sam. “What Is Neural Machine Translation Engine? How Does It Work?” *TranslateFX*, N.p., 18 April 2023. Web. 15 August 2024. <https://www.translatefx.com/blog/what-is-neural-machine-translation-engine-how-does-it-work?lang=en>