

T.C.
İSTANBUL 29 MAYIS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
ÇEVİRİBİLİM ANABİLİM DALI

A STUDY ON CUSTOMIZED NEURAL MACHINE TRANSLATION
EXEMPLIFIED BY IB MYP SUMMATIVE ASSESSMENT CRITERIA

(YÜKSEK LİSANS TEZİ)

Merve ARAZ

Danışman:
Prof. Dr. Işın ÖNER

İSTANBUL
2021

T. C.
İSTANBUL 29 MAYIS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
ÇEVİRİBİLİM ANABİLİM DALI

**A STUDY ON CUSTOMIZED NEURAL MACHINE TRANSLATION
EXEMPLIFIED BY IB MYP SUMMATIVE ASSESSMENT CRITERIA**

(YÜKSEK LİSANS TEZİ)
Merve ARAZ

Danışman:
Prof. Dr. Işın ÖNER

İSTANBUL
2021

T. C.
İSTANBUL 29 MAYIS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Çeviribilim Anabilim Dalı'nda 010518YL05 numaralı Merve ARAZ'ın hazırladığı “*A Study on Customized Neural Machine Translation Exemplified by IB MYP Summative Assessment Criteria*” konulu yüksek lisans tezi ile ilgili tez savunma sınavı, 14/09/2021 günü (17:00 – 18:00) saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin başarılı olduğuna oy birliği ile karar verilmiştir.

Prof. Dr. Işın ÖNER
İstanbul 29 Mayıs Üniversitesi
(Tez Danışmanı ve Sınav Komisyonu Başkanı)

Prof. Dr. Ayşe Banu KARADAĞ
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Nilüfer ALİMEN
İstanbul 29 Mayıs Üniversitesi

BEYAN

Bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadığını beyan ederim.

Merve ARAZ

14/09/2021

ÖZ

Nöral Makine Çevirisi (NMT), çeviri endüstrisinde yeni bir araştırma alanıdır. Nöral Makine Çevirisinin farklı çalışma alanlarında kullanımı konusunda çeşitli çalışmalar yürütülmektedir. Bunlardan biri de eğitim alanıdır. Eğitim, yoğun hacimli veri içeren ve sürekli olarak büyüyen bir alandır. Bu bağlamda Uluslararası Bakalorya (IB) uluslararası arenada kabul görmüş eğitim programlarından biridir. Programın içerisinde sürekli olarak yüksek hacimli çeviri ihtiyacının karşılanması gerekmektedir. Bu nedenle, bu tezde IB'nin yoğun hacimli ve süreklilik teşkil eden çeviri ihtiyaçlarına çözüm sağlamak için özelleştirilmiş nöral makine çevirisi kullanılması amaçlanmıştır. Özelleştirilmiş nöral makine çevirisi sisteminden elde edilen sonuçlar, İstatistiksel Makine Çevirisi (SMT) sonuçları ile karşılaştırılmıştır. Bulgular Çok Boyutlu Kalite Metrikleri (MQM) ve BLUE puanı çerçevesinde değerlendirilmiştir. Son olarak ise yapılan karşılaştırmalar ve analizler doğrultusunda bir sonuca varılmaya çalışılmıştır.

Anahtar Kelimeler:

Makine Çevirisi, Nöral Makine Çevirisi, Özelleştirme, İstatistiksel Makine Çevirisi, Çok Boyutlu Kalite Metrikleri

ABSTRACT

Neural Machine Translation (NMT) is a recent research area in translation industry. Various studies are being conducted on the use of NMT in different fields of study. One of these fields is education. Education is an expanding area with high volumes of data. In this context, the International Baccalaureate (IB) is one of the internationally accepted educational programmes. Great volumes of translation needs are required to be met continuously in the IB programme. Therefore, customized NMT technology is used to provide a solution to high and constant translation needs of the IB in this thesis. The results obtained from the customized NMT system are compared with those of Statistical Machine Translation (SMT). The findings are analyzed within the scope of Multidimensional Quality Metrics (MQM) and BLEU Score metric system. Finally, a conclusion is drawn in line with the comparisons and analyses.

Key Words:

Machine Translation, Neural Machine Translation, Customization, Statistical Machine Translation, Multidimensional Quality Metrics

ACKNOWLEDGEMENTS

Firstly, I would like to offer my thanks to TUBITAK for funding my academic studies with 2210-A General Domestic Graduate Scholarship Program. Its support encouraged me all along my graduate studies.

Above all, I would like to extend my sincere thanks to my esteemed advisor Prof. Işın Öner for her enlightening advice and support with this thesis. She has always been a source of inspiration throughout my academic journey. I have always tried to follow the trail she has blazed.

I would like to express my gratitude to Prof. Ayşe Banu Karadağ for encouraging me to take this exciting road. She has greatly contributed to me all through my undergraduate and graduate education. I also would like to thank Asst. Prof. Nilüfer Alimen for always being available and supportive when I needed her the most.

Lastly, I want to thank my family and friends for always standing by my side and believing in me unwaveringly throughout this journey. Without their encouragement and support, I could not have come this far. I am the luckiest for having you.

TABLE OF CONTENTS

TEZ ONAY SAYFASI	ii
BEYAN	iii
ÖZ	iv
ABSTRACT	v
ACKNOWLEDGEMENTS	iv
LIST OF ABBREVIATIONS	iv
LIST OF FIGURES AND TABLES	iv
INTRODUCTION	1
GENERAL SCOPE	5
1. TRANSLATION TECHNOLOGIES	6
1.1 Computer-Aided Translation	7
1.2 Translation Memory	7
1.3 Termbase	7
1.4 Alignment	8
2. MACHINE TRANSLATION	9
2.1 Rule-Based Machine Translation	9
2.2 Statistical Machine Translation	10
2.3 Neural Machine Translation	10
3. CUSTOMIZATION	12
4. CORPUS	14
4.1 International Baccalaureate	14
4.1.1 International Baccalaureate in Turkey	15
4.2 Middle Years Programme	15
5. DATA PREPARATION	18
5.1 Pre-Alignment	18
5.2 Alignment	19
5.3 Training	22
5.4 Deployment	25
5.5 Using Deployed Model	26
5.5.1 Node.js	27
5.5.2 Visual Studio Code	27

5.5.3	Running Deployed Model	27
6.	MULTIDIMENSIONAL QUALITY METRICS.....	31
7.	CUSTOM TRANSLATOR OUTPUT ANALYSIS	34
7.1	Custom Translator Output Analysis Evaluation.....	42
7.2	BLEU Score of the Customized NMT Outputs.....	42
8.	COMPARISON OF SMT AND NMT OUTPUTS.....	44
8.1	NMT and SMT Output Analysis Evaluation.....	57
8.2	BLUE Score of the Compared NMT and SMT Outputs	57
8.3	Examination on NMT and SMT Outputs for Further Training	60
	CONCLUSION.....	62
	REFERENCES.....	65

LIST OF ABBREVIATIONS

ALPAC:	Automatic Language Processing Advisory Committee
ATL:	Approaches to Learning
BLEU:	Bilingual Evaluation Understudy
CAT:	Computer-Aided Translation
CP:	Career-related Programme
DP:	Diploma Programme
IB:	International Baccalaureate
IDE:	Integrated Development Environment
JRE:	Java Runtime Environment
MT:	Machine Translation
MYP:	Middle Years Programme
MQM:	Multidimensional Quality Metrics
NMT:	Neural Machine Translation
PYP:	Primary Years Programme
QA:	Quality Assessment
RBMT:	Rule-Based Machine Translation
SMT:	Statistical Machine Translation
TM:	Translation Memor

LIST OF FIGURES AND TABLES

Figure 1: Selecting the language pair	20
Figure 2: User interface for editing	21
Figure 3: User interface in case of misaligned segments	21
Figure 4: TMX generation screen on LF Aligner	22
Figure 5: Creating a workspace in Microsoft Custom Translator	23
Figure 6: Creating a project.....	23
Figure 7: Uploading files into the project	24
Figure 8: Creating the model.....	25
Figure 9: Training the model.....	25
Figure 10: Starting the deployment of the model.....	26
Figure 11: Deployment of the model	26
Figure 12: Visual Studio Code Interface.....	29
Figure 13: Visual Studio Code Translation Command	30
Figure 14: Visual Studio Code Translation Output.....	30
Figure 15: MQM Core Representation	31
Table 1	43
Table 2	59
Table 3	60

A STUDY ON CUSTOMIZED NEURAL MACHINE TRANSLATION EXEMPLIFIED BY IB MYP SUMMATIVE ASSESSMENT CRITERIA

INTRODUCTION

Technological developments have affected all business areas in the world and the industry of translation is not an exception. As a result of these technological improvements, Machine Translation (MT) software has been developed. Machine Translation is an automated translation system which is completely carried out by computer software. Thanks to advancements in these systems, they started to provide accurate and usable outputs. Therefore, translators and companies are now able to save time, money and effort and to translate more in less time.

The first machine translation studies began in 1940s, and since then it has developed significantly. There are various types of machine translation systems such as Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT) and Neural Machine Translation (NMT). For several years, SMT has largely dominated the industry. However, being emerged in 2013, NMT has attracted great attention. Currently, it is quite popular in the industry of translation. In addition to the huge interest towards NMT in the sector, there is a great attention in NMT in academy and many NMT studies are being conducted around the world.

NMT has some outstanding characteristics that may surpass SMT in some aspects. The most significant feature of NMT is that it can be specially trained and customized for any domain so long as there is enough data. Thanks to its customizable structure, the output quality of the domain-specific NMT engines is higher in comparison to other generic MT outputs.

NMT systems can be trained for different areas such as legal, medical, finance, technology etc. Although it may not come to one's mind at first, education can also be listed among these fields. The main reason of including education in this list is that it is an ever-growing area. Countless number of documents are produced continuously to teach, test, evaluate, assess and report learners. Especially international education institutions require significant amount of translation in various types, formats and

volume. Lesson contents and plans, unit plans, exams, extracurricular activities, formative and summative assessments, report cards can be listed among the documents that needs to be translated on a weekly, monthly, termly and yearly basis. It should be noted that each of these documents are also separately prepared for different grade levels. Hence, the number of documents may sometimes double, triple or even quadruple. Generally, the time allocated to translation of these documents is usually very short due to the curriculum requirements. Considering all these facts and situations, a customized NMT for education field could be of great use.

As mentioned previously, international education institutions require a great amount of translation need. In Turkey, there are several institutions that provide internationally accredited education programmes. One of the most known and globally accepted education programme that is also implemented in Turkey is the International Baccalaureate (IB). The IB is an education programme that is accepted all over the world. It is designed to provide education for children aged 3-19. The IB offers four different education programmes: Primary Years Programme (PYP), Middle Years Programme (MYP), Diploma Programme (DP), Career-related Programme (CP). More detailed information on the IB education system is given in the fourth chapter.

Currently, there are over 90 IB World Schools that provide IB education in Turkey¹. Although programmes are intertwined and complementary, each has its own specific content and materials. This results in a voluminous amount of translation need to be met. Hence, a custom NMT engine can provide faster, accurate and consistent translations to meet this demand. To this end, each IB programme is examined to find the most suitable programme for which a customized NMT can be created. As a result of these examinations, it is found that MYP provides the most appropriate corpus to generate a customizable NMT.

MYP programme is designed for students aged 11 to 16. It is a framework that focuses on how to teach and assess and not on what to teach. It covers the following

¹ “Find an IB World School,” International Baccalaureate Organization, accessed August 30, 2021. <https://ibo.org/programmes/find-an-ib-school/?SearchFields.Region=&SearchFields.Country=TR&SearchFields.Keywords=&SearchFields.BoardingFacilities=&SearchFields.SchoolGender=&page=4>

subject groups: Arts, Design, Individuals and Societies², Language Acquisition, Language and Literature, Mathematics, MYP Projects³, Physical and Health Education, Science. In Turkey, MYP includes students in 5th grade to 9th grade since students belonging these grade levels are in the range of MYP age group.

All the students between these grade levels receive the abovementioned lessons within the scope of MYP. To pass the class, students need to accomplish specified tasks. One of these is the assessment criteria. In MYP, assessment criteria are divided into two: formative and summative. Formative assessment criteria comprise of a little part of year-end score while summative assessment criteria have a great role in determining the overall score of students. In addition, summative assessment criteria have a technical structure unlike formative assessment criteria which do not have any technical structure and can be designed in accordance with the needs of the course. Students are responsible to deliver four summative assessment criteria for each lesson in a term. The grades that students receive from these criteria significantly affect their MYP scores. Throughout the year, students receive several summative assessment criteria that results in a continuous and voluminous translation need.

To this end, the aim of this thesis is to customize an NMT engine in order to save time and effort in the translation process of MYP summative assessment criteria. In addition, more accurate, consistent and quality outputs is aimed to obtain through the customized engine. In this context, another aim of this thesis is to compare the quality of customized NMT output with the generic SMT output in order to observe the quality difference between the SMT and NMT software. In the process of comparing the quality of the translated texts, BLEU Score and Multidimensional Quality Metrics (MQM), which will be explained in the relevant chapters, will be applied.

In order to accomplish this goal, an NMT system will be trained with all the available summative assessment criteria. Later, the system will be tested with actual summative assessment criteria i.e. documents that are prepared for actual lessons and assigned to students in an IB World School. The findings will be compared with the

² Individuals and Societies may cover more than one lesson in accordance with the Ministry of Education of the country where it is implemented.

³ MYP Projects is not a lesson but a long-term project that students need to conduct in the final year of the MYP to complete the program.

outputs of other SMT engines to examine the quality difference. The comparison will be held in accordance with the framework of the Quality Assessment (QA) metrics. At the end of the study, concluding remarks will be provided.

It is important to note that since the summative assessments are licensed publishing of the IB, they will not be shared explicitly at the appendices of this study due to copyright issues.

Several studies have been conducted on NMT and customization so far. However, there is not a study that focuses on the translation of the IB MYP summative assessment criteria through a customized NMT engine. Hence, it is hoped that this study would be helpful for educators and translators working in this field and more research would be carried out in the future.

GENERAL SCOPE

In the first chapter, a basic framework of Computer-Aided Translation (CAT) tools is presented. Relevant concepts such as Translation Memory (TM), Termbase and Alignment are explained within the scope of this chapter.

In the second chapter, different machine translation systems such as RBMT, SMT and NMT are described in detail to provide basic information about machine translation.

In the third chapter, the concept of customization is defined. The benefits of using a custom engine are clarified. In addition, it is stated that in which cases it is more beneficial to use a custom engine.

In the fourth chapter, the background of the corpus of the research is explained. An analysis of the corpus is also provided in this chapter.

In the fifth chapter, the data preparation of the study is explained in detail. Pre-alignment, alignment, training, deployment and running the deployed model processes are specified respectively.

In the sixth chapter, Multidimensional Quality Metrics are described and the specific metrics that are used in the current study are designated to assess the quality of translations.

In the seventh chapter, some examples of NMT outputs are shared. In addition, the BLEU score information of the deployed model is provided within the scope of this chapter.

In the eighth chapter, various generic SMT outputs are compared with the customized NMT engine outputs in order to analyze the quality difference. MQM findings are also given as to the qualities of translated texts.

In the last chapter, previous sections are briefly summarized and concluding remarks are provided together with recommendations for future studies.

1. TRANSLATION TECHNOLOGIES

It would not be wrong to claim that the translation has a long history. Although there is not a specified period of time that points out the first translation made in the history, it can be assumed that translation exists since the existence of language itself. However, despite having a deep-rooted history, translation proved itself as a scientific and separate discipline only at the end of 1970s. Being accepted in the academic circles, it received the name translation studies which was designated by James S. Holmes after his famous article titled “*The Name and Nature of Translation Studies*” (Holmes 1988, 71). Holmes also indicated the main areas of study within the discipline. Upon this article, Gideon Toury published another study which demonstrated the Holmes’ map of translation and explained it (see Toury 1995, 10). In accordance with the map, various sub-fields were developed such as theoretical translation studies, descriptive translation studies and applied translation studies (cf. Toury 1995). As time has passed and technology has progressed, new sub-fields have developed in line with the needs of the period. One of these fields is translation technologies.

It is a fact that technology has rapidly developed and affected every business area. Over the last thirty years, technological developments have had significant effects on translation industry. With the increase in computerization, various tools and software that facilitate the translation process have emerged. Some of these products are computer-aided translation tools, translation memory, termbase and alignment tools. The advent of these tools contributed to an increased productivity, a better quality and more consistency in the translation business (cf. Doherty 2016, 962).

Another reason why translation tools have importance is because they allow practitioners to translate, edit and store source and target texts. Thanks to these features, translation tools facilitate consistency, accuracy and increased speed during translation process. Apart from computer-aided translation tools, different tools have been created to provide benefit to translators such as project management tools that allow several translators to work collaboratively and terminology management tools which enable organizing terms to acquire accurate translations. In the following part, definitions and

scope of these tools is given in order to provide a framework for the concepts used in the translation technologies.

1.1 Computer-Aided Translation

Computer-Aided Translation (CAT) system is software that is mainly used by translators. A CAT tool is basically composed of Translation Memory, Termbase, Dictionary and Machine Translation engines⁴. Currently, the use of CAT tools is a must for translators and linguists since it immensely supports the translation process.

To begin with the benefits of a CAT tool, firstly it can be said that it facilitates translating process thanks to its user-friendly interface. The user interface of CAT tools is generally divided into two main sections, one for source text segments and one for target text segments. Depending on the tool, additional areas to show terms and matches are also found on that page. Secondly, it enables editing of translations. Making additions, omissions or any kind of editing becomes easier to carry out on a CAT tool programme. Last but not least, CAT tools enable the storing and archiving of source and target texts thanks to translation memories. This feature allows users to easily access previous translations when needed.

1.2 Translation Memory

Translation Memory is a database that stores source and target sentences of a previously translated text for future use. Thus, translators are able to reuse formerly translated segments. Since previous translations become always available to translators thanks to translation memories, the accuracy and consistency of translations substantially increase. Moreover, translation memories also improve the consistency of translations which directly enhance the translation quality. Another advantage of translation memory is that they increase the speed of translation process.

1.3 Termbase

Termbase is another unit of a CAT tool. It is similar to a dictionary. However, it includes terms and expressions that are specific for an area, client and/or product. The benefits of using termbase are being able to produce consistent translation, having a higher and more accurate output quality and experiencing a speedy translation process. Termbase files are

⁴ “What Is A CAT Tool?,” Trados, accessed April 4, 2021. <https://www.sdltrados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html>

produced in a couple of common file formats such as .XDT, XML, .TMX which enables translators to integrate termbase files into their own software⁵.

1.4 Alignment

Alignment software is generally an inseparable part of CAT tools. Basically, it is software that enables the pairing and aligning of two and more documents. In addition to being integrated into CAT tool systems, there are stand-alone alignment tools, as well.

This software is used in translation industry to use previously translated texts. Thanks to alignment tools, source and target texts can be aligned and converted into different file types such as TMX or ALIGN. Upon the completion of the alignment process, aligned files can be used in different tools and software including CAT tools and the training of customizable MT engines.

Upon the descriptions of the tools and software that are mostly used in CAT tools and translation industry, another pillar of translation technologies comes to the front to be addressed: machine translation.

The latest developments in the technology and translation sector have had remarkable effects on machine translation. It can be claimed that human and computer cooperation is observed in translation sector recently. Over time, translation has become a technology-dependent industry. The following chapter focuses on the history and branches of machine translation and provides information about the current situation of the area.

⁵ Christopher Dell, “Why Use a Termbase or a Glossary When Translating an App?”, Locale, accessed August 30, 2021. <https://www.locale.to/why-use-a-termbase-or-glossary-when-translating-an-app/>

2. MACHINE TRANSLATION

The history of machine translation goes back to 1950s. In 1949, Warren Weaver who was of Rockefeller Foundation argued the possibility of machine translation for the first time. His proposal became widespread and several machine translation researches began in US universities. In 1964, Automatic Language Processing Advisory Committee (ALPAC) was established (Hutchins 1995, 436). In 1966, ALPAC published a report against machine translation studies. In consequence, most of the machine translation funding stopped in the US for almost a decade. However, there were some studies that continued meanwhile. In 1980s, SYSTRAN and Logos became prominent with the machine translation engines that they produced. These studies were based on linguistic rules. Only after 1990s, statistical methods began to be used. Statistical machine translation studies created a great interest towards MT. In the following years, new phrase-based approaches were also developed. However, the ground-breaking studies were carried out in neural machine translation. NMT was first proposed by Kalchbrenner and Blunsom in 2013. Today, it is the state-of-the-art MT system. In the below sections, descriptions, advantages, and disadvantages of Rule-Based Machine Translation, Statistical Machine Translation and Neural Machine Translation are discussed.

2.1 Rule-Based Machine Translation

Rule-Based Machine Translation is based on huge sets of linguistic rules and bilingual dictionaries for each language. While RBMT requires great amount of linguistic rules, it does not necessitate parallel texts. Although not requiring huge parallel texts might seem as an advantage of RBMT at first, continuously writing and editing rules and referring to bilingual dictionaries is a compelling disadvantage due to the effort this process requires.

It should also be noted that there are diverse types of RBMT systems and they can generate quality output. However, there are some setbacks to this system. First of all, because of the richness of language, new words and phrases are added to both source and target languages. Therefore, RBMT systems need to be maintained and updated constantly. Secondly, the linguistic rules are written and developed by human linguists and software developers. Considering these, it can be said that RBMT is a relatively expensive and time-consuming system.

2.2 Statistical Machine Translation

The *modus operandi* of Statistical Machine Translation differ from that of RBMT. While RBMT is based on linguistic rules and bilingual dictionaries, SMT mainly relies on parallel corpora which can be obtained from translation memories and glossaries. The level of parallelism of the corpora to be used needs to be on the sentence level.

SMT has some advantages when compared with RBMT systems. For instance, since RBMT entails human effort it is an expensive system in comparison to SMT. Furthermore, SMT can provide more accurate outputs if trained with domain-specific corpora. Yet, SMT also has some disadvantageous features. To give an example, creating SMT system requires some degree of computational skills. Moreover, the corpora of SMT systems are obtained through the Internet regardless of the quality of the texts. This may sometimes result in a reduced translation quality which leads to inaccurate and inconsistent outputs.

2.3 Neural Machine Translation

Neural Machine Translation is a corpus-based machine translation (Forcada 2017, 292). According to the Kyunghyun Cho (Cho et al. 2014, 103), NMT is a new approach based on pure neural networks (i.e. *artificial neural networks*). It needs to be trained on huge corpora in other words source and target language segments that are obtained from extensive translation memories. It works on the basis of encoder-decoder pair. NMT uses neural networks to translate entire sentences without breaking them down into smaller parts⁶. Hence, it puts forth higher quality translations.

One of the advantage of NMT is that it is able to minimize the syntax-related translation mistakes. Unlike RBMT or SMT systems, NMT can be “directly trained both source and target language”⁷.

As mentioned, NMT systems require a training process to operate. The term used for this process is called “customization”. For NMT systems that work with neural networks, they require to be customized with specific corpora to provide usable material.

⁶ Daniel Brockmann, “SDL Machine Translation: The Future of Neural Machine Translation Is Here”, Trados, accessed April 4, 2021. <https://www.sdltrados.com/blog/future-nmt.html>

⁷ “Neural Machine Translation: Embarking on the New Frontier”, Ulatius, accessed September 1, 2021. <https://www.ulatus.com/translation-blog/neural-machine-translation-embarking-on-the-new-frontier/>

In other words, customization carries an important role. Hence, the following chapter concentrates on customization.



3. CUSTOMIZATION

Machine translation greatly facilitates the translation process in several aspects. It is a fact that generic machine translation systems are able to provide understandable results. However, when it comes to the translation of industry-specific texts, some difficulties arise in the production of publishable output. They remain insufficient in domain-specific syntax and language use, terminology, accuracy and consistency. However, when a customized NMT in other words an engine that is trained with domain-specific data is used, the quality of the output increases. These engines reduce the probability of experiencing previously mentioned problems as well as diminishing the human effort in post-editing process.

Customization can be described as the process in which a neural machine translation system is trained with area-specific documents such as previous translations, glossaries, terminology. Training of an engine is realized by uploading the necessary files to the system. It should also be noted that training process might take some time depending on the volume of the data and the system that is preferred to be used.

The cutting-edge neural machine translation systems can be customized in line with the needs of the domain. To this end, the main benefit of custom NMT is higher translation accuracy. When an engine is trained with area-specific and quality data, it produces more accurate target texts. Furthermore, thanks to the flexibility of software, the engine can be constantly trained with more volumes of data. This is also another factor to increase the translation quality in the long term. Another benefit of customized neural machine translation is that the engine learns the translation style and creates more consistent outputs.

Customized neural machine translation is suitable for any field with high volume of texts, specific document type or continuous translation needs. In addition, custom NMT engines are able to produce more quality output in technical texts. Hence, before deciding whether to use custom NMT or generic MT, general features of the domain, text types and structure, and translation demand should be analyzed in detail. In light of this information, a decision should be made on the machine translation engine. Subsequently, a corpus needs to be formed.

As a consequence, a framework will be drawn for the corpus of the current thesis and it will be created in line with it. Detailed information on corpus is provided in the fourth chapter. The properties of the corpus, technical features about its structure is shared. In addition, background information related to the corpus content is given in the following section.



4. CORPUS

A corpus is necessary in order to customize a machine translation system, especially a neural machine translation engine. In this context, a corpus may include parallel texts, terminologies, glossaries, bilingual dictionaries, etc. Generally, to train a neural machine translation engine, thousands of sentences may be required.

For the purposes of this thesis, a corpus is needed to be constructed to train the neural machine translation engine. More information will be given on the contents of the corpus in question in the Middle Years Programme subheading of this chapter. However, before delving into the details of the corpus, the International Baccalaureate, the International Baccalaureate in Turkey and the Middle Years Programme will be explained respectively.

4.1 International Baccalaureate

The International Baccalaureate is an education system that offers challenging and quality programmes to students aged 3-19⁸. To this end, the IB designed special education programmes that are internationally applied and accepted. The diploma awarded by the IB is recognized by the world's best universities and colleges. The success behind the IB programme lies within its education system. To implement an IB programme, schools must be authorized by the International Baccalaureate Organization. Any school that has successfully completed the authorization process is called an IB World School⁹. And schools that offer at least three of the main IB programmes are called IB Continuum School.

As mentioned in earlier the introduction chapter, the IB offers four education programmes: Primary Years Programme (PYP), Middle Years Programme (MYP), Diploma Programme (DP) and Career-related Programme (CP). Currently, IB programmes are offered only in English, French and Spanish¹⁰. Each programme has specified subject groups and a special curriculum. The IB education is flexible to embrace the local curricula of the state in which it is applied. It only offers a framework on methods

⁸ "About the IB", International Baccalaureate Organization, accessed April 2, 2021. <https://www.ibo.org/about-the-ib/>

⁹ "Programmes", International Baccalaureate Organization, accessed April 12, 2021. <https://www.ibo.org/programmes/>

¹⁰ "Teaching the IB in Your Language", International Baccalaureate Organization, accessed April 16, 2021. <https://www.ibo.org/benefits/teaching-ib-in-your-language/>

of teaching, it does not decide what to be taught to learners. In this regard, it can be claimed that it has an adaptable structure. This allows the IB to be applied all over the world regardless of the different structure of the education system where it is implemented. And Turkey is one of the countries in which the IB is applied commonly.

4.1.1 International Baccalaureate in Turkey

The first IB World School started to operate in 1994 in Turkey. Currently, there are 92 IB World Schools that offer one or more of four programmes in the country¹¹. In Turkey, PYP is offered by 45 schools, MYP by 12 schools and DP by 58 schools¹².

In this thesis, the MYP is selected to be studied among all four programmes. There are several reasons behind this decision. First of all, it is found that a criteria-based assessment is not available in PYP. Due to the young age-level of PYP learners, different assessment solutions are offered in PYP system. Secondly, DP is completely implemented in English in Turkey. It means that there is not a consistent translation demand for this programme. Thirdly, CP is not offered in any school in Turkey at present. Therefore, there is no translation need for CP currently. However, MYP is offered both in Turkish and English in Turkey. Some schools apply only a couple of lessons in English in MYP, while the rest of the lessons are held in Turkish. In addition, MYP includes a criteria based assessment system which has a significant role in the MYP. These criteria have a specific and stable structure that can be translated through machine translation systems. Considering all these facts, MYP is the only programme that needs be studied for the purposes of this thesis. Details of the MYP, the MYP criteria system will be provided in the following subheading as well as with the information on the corpus of this thesis.

4.2 Middle Years Programme

It is designed for students aged 11-16. In other words, it may encompass 5th to 9th grade students. IB Schools are able to apply the MYP system for three, four or five years in accordance with their education plan.

As explained previously, the MYP is a framework that specifies how to teach, assess and report. As for the teaching system of the MYP, it can be stated that the MYP

¹¹ “Turkey”, International Baccalaureate Organization, accessed April 12, 2021.

<https://www.ibo.org/about-the-ib/the-ib-by-country/t/turkey/>

¹² “Turkey”. International Baccalaureate Organization.

has peculiar subject groups. These subject groups are as follows: Arts, Design, Individuals and Societies, Language Acquisition, Language and Literature, Mathematics, MYP Projects, Physical and Health Education, Sciences. When it comes to the assessment system of the MYP, it has two types of assessment criteria: formative and summative assessment. These assessment criteria are assigned to the students to evaluate and assess their levels. Formative assessments are smaller tasks that are appointed during the process of learning. They do not have a specific text structure. Formative assessments can be prepared in line with the needs of the course. Summative assessments are more comprehensive and may include exams, projects and essays¹³. The most important feature of summative assessment criteria is that they have a specific text structure that do not change by the requirements of the course. In case that different tasks to be assigned to the students, they are added to the summative task without changing the structure of the criteria. Unlike formative assessments, summative assessment criteria have a greater role in determining the overall achievements of the students. The total score students obtain from the summative assessments mostly determine their year-end score.

It should also be noted that the MYP is divided into three year-groups in terms of grade-levels: Year 1, Year 3, Year 5. In other words, Year 1 refers to 5th and 6th grades, Year 3 refers to 7th and 8th grades and Year 5 refers to 9th grade. In the MYP, each subject group has its own summative assessment criteria that are determined by the IB in detail. And each summative assessment contains four criteria: Criterion A, Criterion B, Criterion C and Criterion D. Since the names of the criteria differ in accordance with the subject group, a generic name (i.e. Criterion A) is preferred to define them. Only Language Acquisition covers six criteria and MYP Projects include two criteria. In fact, each year group includes four criteria. For example, Mathematics subject group has four criteria for Year 1, another four criteria for Year 3, and another four criteria for Year 5. That makes a total of twelve criteria for one lesson. However, since only one year-level is applied for a grade-level, i.e. Year 5 for 9th grade, students are responsible for four criteria for one lesson.

Since students are assigned to four criteria for one lesson in a term, as this number is multiplied with the number of lessons, i.e. up to eleven and/or twelve lessons, total

¹³ “MYP Assessment”, Istanbul International Community School, accessed May 23, 2021. <https://www.iics.k12.tr/academics/myp-assessment/>

number of the assigned summative assessment criteria in a year turns out to be quite high. In other words, in a case where all summative assessment criteria need to be translated, the workload becomes tremendous due to time constraints of the curricula.

Apart from the summative assessment criteria, there are also Approaches to Learning (ATL) Skills in MYP. In fact, the ATL Skills are not peculiar to the MYP. They are dealt meticulously in all IB programmes. The main objective of ATL Skills is to enable students learn how to learn. ATL Skills are comprised of five main groups: Thinking, Self-management, Communication, Social, Research. These groups are also divided into several sub-groups. ATL Skills are integrated into all lessons and assessments. Each summative assessment criterion must focus on one or two ATL Skill(s) that is aimed to be internalized by the student.

When creating the corpus of this thesis to customize the NMT engine, summative assessment criteria (both Turkish and English) and the ATL Skills are included. These documents are obtained from the updated IB MYP booklets that are downloaded through the My IB account that is possessed by me. However, previous versions of the IB MYP booklets can be found on the Internet. These booklets are also included in the corpus. In the booklets, aims, objectives, planning, teaching and assessment curriculum of the lessons are described meticulously. As the files are determined, necessary steps to prepare data for the customization process of the NMT engine are taken. These steps can be listed respectively as follows: pre-alignment, alignment, deployment and running the deployed engine. Detailed information on the data preparation process is provided in the chapter below.

5. DATA PREPARATION

In order to train and customize the NMT engine, data that comprises the corpus of this study is ascertained in the previous chapter. This data includes English and Turkish summative assessment criteria, ATL Skills sheet and MYP lesson booklets. Information on the volume of this data should be provided to better understand the scope of the corpus. In total, 116 summative assessment criteria both in English and Turkish are used. It amounts to 63,827 words in all. As for the ATL Skills sheet, it consists 15,078 words. In addition to these documents, booklets are also used to train the engine.

Before starting to the training process, obtained data should be aligned. However, all of the documents should be prepared for the alignment. This is called the pre-alignment process. It is explained in detail in the following section.

5.1 Pre-Alignment

Before starting the alignment process, a couple of pre-alignment steps are taken. First of all, files are divided into two categories: to be edited vs. not to be edited. Summative assessment criteria and ATL Skills sheet fall into the “to be edited” category while lesson booklets are dealt within the “not to be edited” category.

Files in the first category are going to be uploaded to the alignment tool while the ones in the second category are to be directly uploaded to the engine. The reason why not uploading the files in the first category to the engine without aligning them is that they are great in number, i.e. 116 documents, and it may take too much time to add them one by one. By the alignment process, all of the aligned files could be uploaded into the system at one time. Another reason is that each document in the second category is voluminous to upload to the alignment tool and compare any misaligned line.

As to the “to be edited” category, English and Turkish summative assessment criteria are classified into year-groups such as “Individuals and Societies – Year 1”, “Individuals and Societies – Year 3”, “Individuals and Societies – Year 5”, “*Bireyler ve Toplumlar – Yıl 1*”, “*Bireyler ve Toplumlar – Yıl 3*”, “*Bireyler ve Toplumlar – Yıl 5*”. After the completion of year groups, each document is examined regarding its stylistic features. The criteria that are obtained through different file formats had different fonts, font sizes and tables. Since different stylistic features could have resulted in misaligned sentences due to different placement of sentences, the fonts, font sizes and table design

of all the summative assessment criteria set to the same format: Times New Roman, 12 pt. Tables are prepared on MS Word by doing the following steps: Insert > Table > Insert Table. Having designed the above-mentioned properties of the documents, pre-alignment steps of the first category is completed.

As for the “not to be edited” category, the documents are converted into the same file format since the NMT engine does not accept the upload of the different file formats of the same document. For example, if Design booklet is in the PDF file format and *Tasarım* booklet (Turkish of the Design booklet) is in the MS Word file format, the NMT engine would fail in the upload process of these documents in the training process. Therefore, two files should be in the same file format. In this case, *Tasarım* booklet is converted into a PDF file. In so doing, files can be uploaded to the engine successfully. Similar process is run for the rest of the booklets.

Apart from these steps, no other editing is conducted on the documents. Especially, contents of files are protected since they are original IB documents that are used in all IB contexts.

5.2 Alignment

In order to train a neural machine engine, files are required to be aligned and converted into a suitable file format that the software supports. To this end, several alignment tools are available. Some of these are AlignFactory Desktop, Matecat, LF Aligner, SDL Align, Stingray and YouAlign. All of these are stand-alone alignment tools. In the current study, LF Aligner is preferred since it is lightweight, easy-to-use and has a user-friendly interface.

When starting the alignment process in the LF Aligner, firstly the language of source and target files are selected on the screen that pops up.

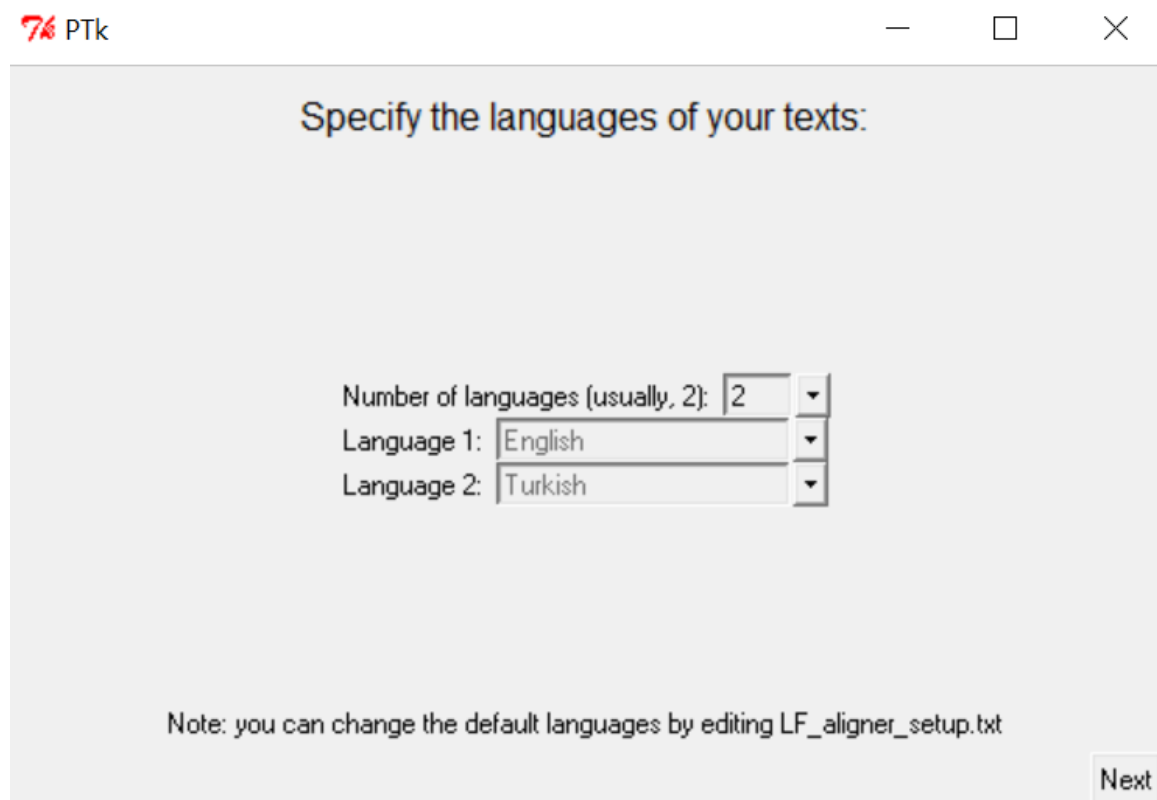


Figure 1: Selecting the language pair

Then, pairing files are uploaded to the LF Aligner, i.e. “Arts – Year 1 and *Sanat – Yıl 1*”. Through the user interface of the LF Aligner, source and target files are examined to check if there are any misaligned segments. If the segments are correctly aligned, the screen appears as follows:

74 LF Alignment Editor 1.8 - aligned_MYP_Year1_Arts_Cr.A_B_C_D-MYP_Y11_Sanat_Cr.A_B_C_D.txt

1	Criterion A: Knowing and understanding	Kriter A: Bilme ve anlama	MYP_Year1_Arts_Cr.A_B_C_D
2	Maximum: 8	Maksimum: 8	MYP_Year1_Arts_Cr.A_B_C_D
3	At the end of year 1, students should be able to:	1.yılın sonunda öğrencilerin aşağıdakileri gerçekleştirebiliyor olmaları gerekir:	MYP_Year1_Arts_Cr.A_B_C_D
4	* demonstrate awareness of the art form studied, including the use of appropriate language	* Uygun dilin kullanımıyla birlikte, işlenen sanat dalındaki farkındalığını göstermek	MYP_Year1_Arts_Cr.A_B_C_D
5	* demonstrate awareness of the relationship between the art form and its context	* Sanat türü ve bağlamı arasındaki ilişkideki farkındalığını göstermek	MYP_Year1_Arts_Cr.A_B_C_D
6	* demonstrate awareness of the links between the knowledge acquired and artwork created.	* Edinilen bilgiyle yaratılan sanat eseri arasındaki bağlantılara dair farkındalığını göstermek	MYP_Year1_Arts_Cr.A_B_C_D
7	Achievement level	Başarı düzeyi	MYP_Year1_Arts_Cr.A_B_C_D
8	Level descriptor	Düzy tanımlayıcısı	MYP_Year1_Arts_Cr.A_B_C_D
9	0	0	MYP_Year1_Arts_Cr.A_B_C_D
10	The student does not reach a standard described by any of the descriptors below.	Öğrenci aşağıdaki tanımlayıcılardan hiçbirinin standardına erişememiştir.	MYP_Year1_Arts_Cr.A_B_C_D
11	1-2	1-2	MYP_Year1_Arts_Cr.A_B_C_D
12	The student:	Öğrenci:	MYP_Year1_Arts_Cr.A_B_C_D
13	* demonstrates limited awareness of the art form studied, including limited use of appropriate language	* Uygun dilin sınırlı olarak kullanımıyla birlikte, işlenen sanat dalında sınırlı farkındalık gösterir.	MYP_Year1_Arts_Cr.A_B_C_D

Merge (F1) Split (F2) Shift up (F3) Shift down (F4)

Figure 2: User interface for editing

However, if there are misaligned segments, the screen is seen as in the figure below:

74 LF Alignment Editor 1.8 - aligned_MYP_Year3_Mathematics_Cr.A_B_C_D-MYP_Y13_Matematik_Cr.A_B_C_D.txt

30	* generally solve these problems correctly.	* bu problemleri çeşitli bağlamlarda genellikle doğru bir biçimde çözebiliyor.	MYP_Year3_Mathematics_Cr.A_B_C_D
31	Criterion B: Investigating patterns	Kriter B: Örüntüleri inceleme	MYP_Year3_Mathematics_Cr.A_B_C_D
32	Maximum: 8	Maksimum: 8	MYP_Year3_Mathematics_Cr.A_B_C_D
33	At the end of year 3, students should be able to:	3.yılın sonunda öğrenciler,	MYP_Year3_Mathematics_Cr.A_B_C_D
34	* select and apply mathematical problem-solving techniques to discover complex patterns	* matematiksel sorun çözme tekniklerini karmaşık örüntüleri keşfetmek için seçebilmeli ve uygulayabilmeli	MYP_Year3_Mathematics_Cr.A_B_C_D
35	* describe patterns as relationships and/or general rules consistent with findings	* Örüntüleri ilişkiler ve/ya bulgularla tutarlı genel kurallar olarak betimleyebilmeli	MYP_Year3_Mathematics_Cr.A_B_C_D
36	* verify and justify relationships and/or general rules.	* ilişkileri ve/ya genel kuralları doğrulayabilmeli ve gerekçelendirebilmeli.	MYP_Year3_Mathematics_Cr.A_B_C_D
37	Achievement level	Başarı düzeyi	MYP_Year3_Mathematics_Cr.A_B_C_D
38	Level descriptor	Düzy tanımlayıcısı	MYP_Year3_Mathematics_Cr.A_B_C_D
39			MYP_Year3_Mathematics_Cr.A_B_C_D
40	The student does not reach a standard described by any of the descriptors below.		MYP_Year3_Mathematics_Cr.A_B_C_D
41	1-2	0	MYP_Year3_Mathematics_Cr.A_B_C_D
42	The student is able to:	Öğrenci aşağıdaki tanımlayıcıların herhangi birinde betimlenen bir	MYP_Year3_Mathematics_Cr.A_B_C_D

Merge (F1) Split (F2) Shift up (F3) Shift down (F4)

Figure 3: User interface in case of misaligned segments

In such a case, the misaligned segments are detected firstly. Later, necessary fixation is made. For example, if the segments are merged they are split using F2 key. If a counterpart of a sentence is shown one or two segments below than where should be (See Figure 3), the misaligned segments are shifted up using F3 key. After all the misalignments are fixed manually by me, the file is saved and the output is received

through the following steps: File > Save & Exit. At the final stage, another screen pops up:

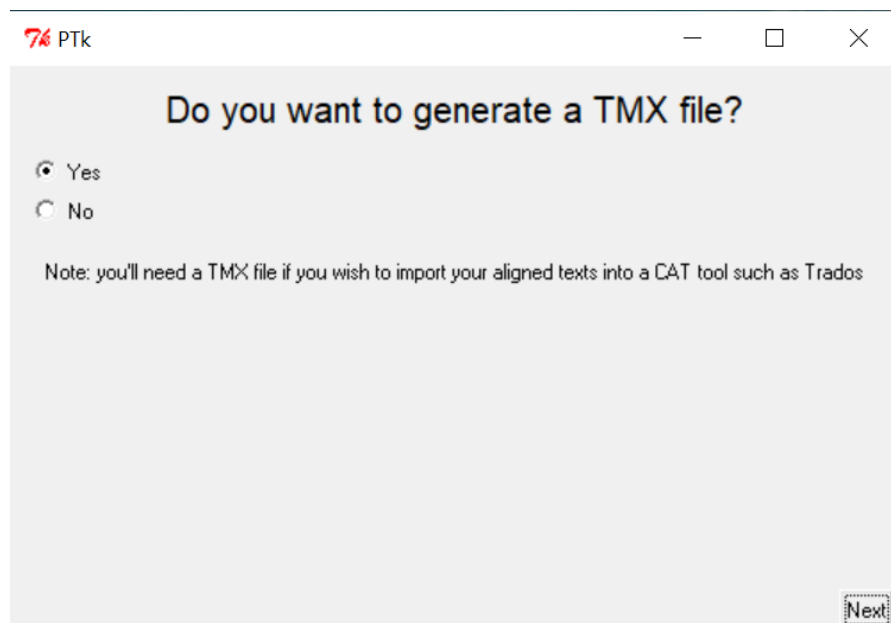


Figure 4: TMX generation screen on LF Aligner

On this screen, the “Yes” option is selected and the TMX file that is needed for the training of the engine is obtained. These steps are repeated for all the documents to be aligned.

If misaligned segments are not corrected, the engine would be trained with mistranslations. As in the example in Figure 3, segments 37 and 38 which are “Achievement Level” and “Level Descriptor” phrases would not be translated as they should be while segments 38 and 39 that are “0” and “The student does not reach a standard described by any of the descriptors below.” would be translated as *Başarı Düzeyi* and *Düzey Tanımlayıcısı*, respectively. Similar mistakes could be encountered if not detected and fixed.

5.3 Training

There are several neural machine translation systems that allow users to create customizable models. Amazon Translate, Google AutoML Translation, KantanMT, Microsoft Custom Translator, Systran and Yandex Translate can be listed among major cloud-based NMT system providers.

While Google's NMT system currently supports only TVS and TMX file formats, Microsoft Custom Translator supports several other file formats such as TXT, HTML, HTM, PDF, DOCX, ALIGN, TMX, XLF, XLIFF, LCL, XLSX, ZIP. Taking this difference into consideration, Microsoft Custom Translator is found more suitable for the purposes of this thesis.

Having decided on the NMT system, setting up and creating workspace processes are carried out (see Figure 5).



Figure 5: Creating a workspace in Microsoft Custom Translator

Once a workspace is created, a project is generated within the workspace.

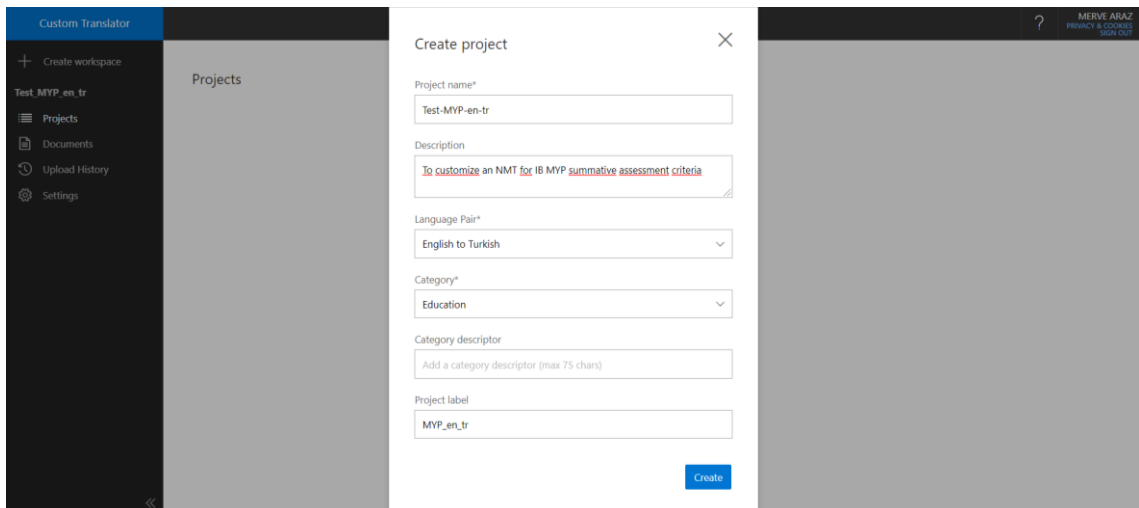


Figure 6: Creating a project

The documents previously aligned on LF Aligner are uploaded into the project. This process could take a while depending on the volume of files. In this stage, 116 summative assessment criteria are uploaded to the system. This amount of criteria corresponds to 63,827 words in total in both languages. Furthermore, ATL Skills table which equals to 15,078 words and MYP lesson guide booklets are also uploaded to the engine (see Figure 7).

In order to train a model, a minimum of 10,000 parallel sentences are required for Microsoft Custom Translator engine¹⁴. With the upload of all these documents, the minimum requirement of Microsoft Custom Translator is met. In fact, as long as this requirement is not met, the Custom Translator does not enable the start of training process.

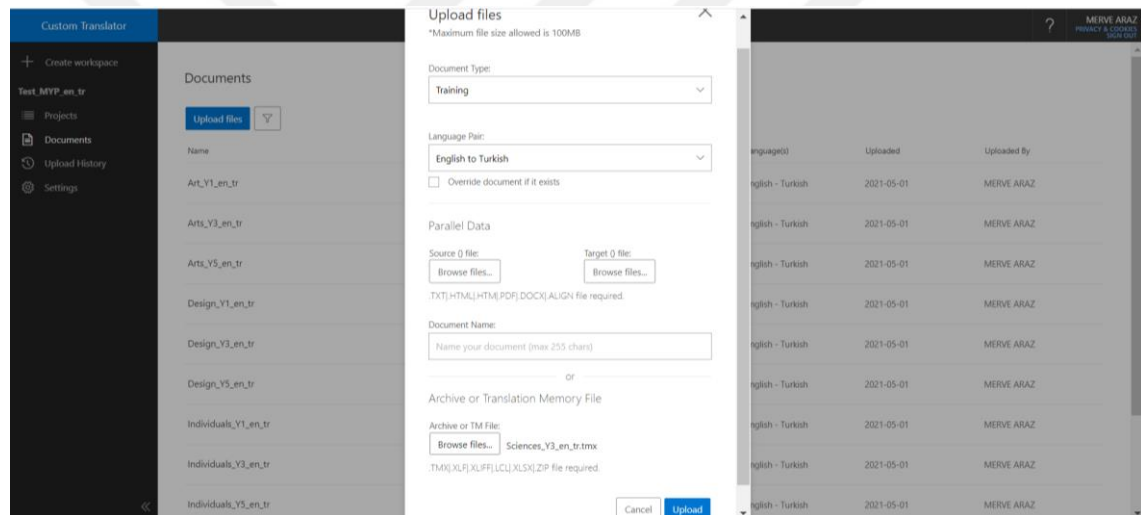


Figure 7: Uploading files into the project

A total of 21,700 English and Turkish sentences are uploaded to the customized model and the system used 5,611 training sentences. Upon the completion of the upload of the files, the model is created. Afterwards, deployment of the model is realized for the model to be used.

¹⁴ “How to Train a Model”, Azure Cognitive Services, accessed February 28, 2021. <https://docs.microsoft.com/en-us/azure/cognitive-services/translator/custom-translator/how-to-train-model>

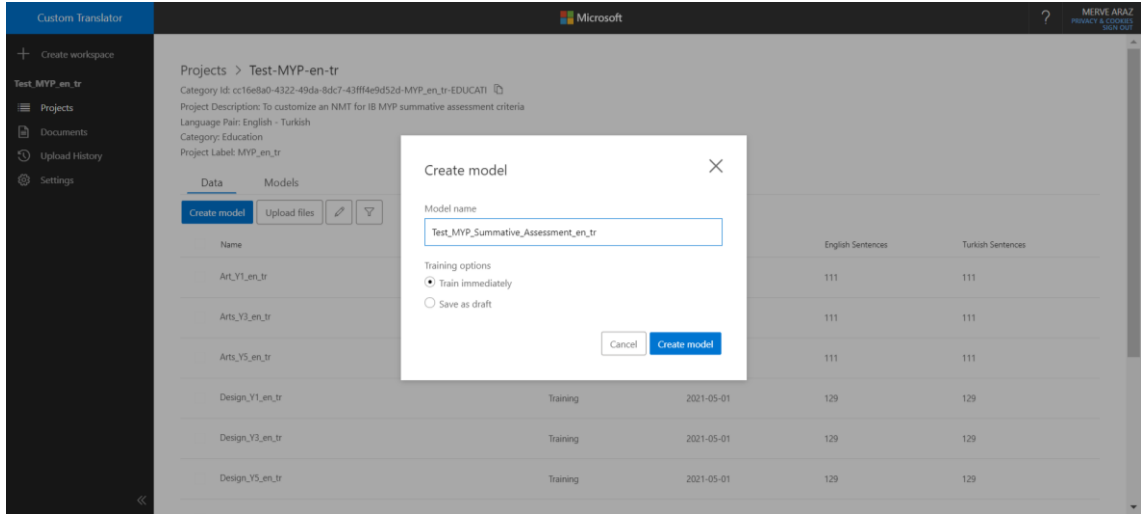


Figure 8: Creating the model

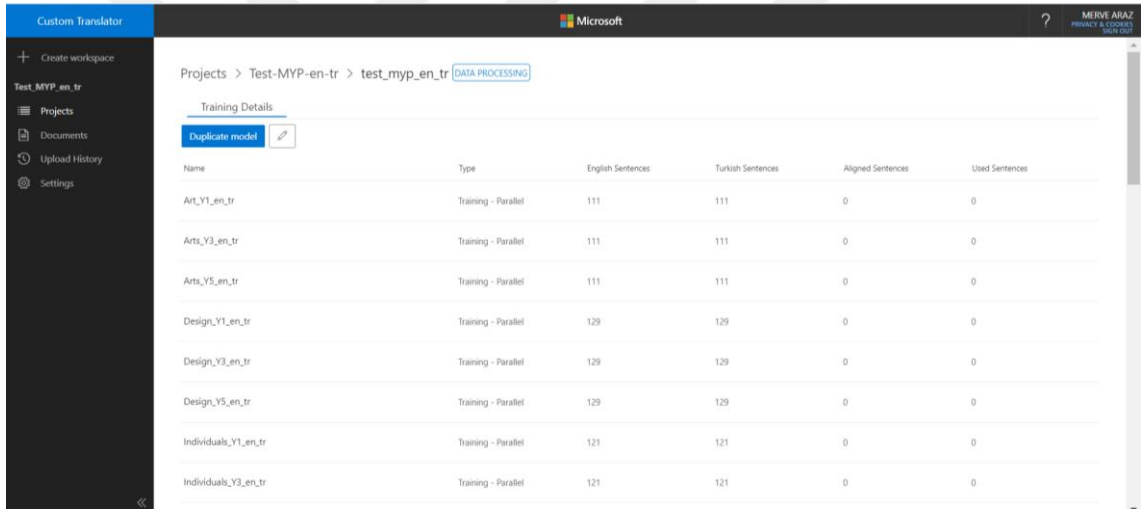


Figure 9: Training the model

5.4 Deployment

Before the deployment of the model, it should be ensured that all of the required training data is uploaded into the model. When it comes to deployment process, deployment region must be selected (see Figure 10). In Microsoft Custom Translator there are three regions: North America, Europe and Asia Pacific. In the model used in this thesis, Europe region is selected. Then, the deployment process is started by clicking the save and deploy buttons. As the deployment process is finished, model status is changed as deployed.

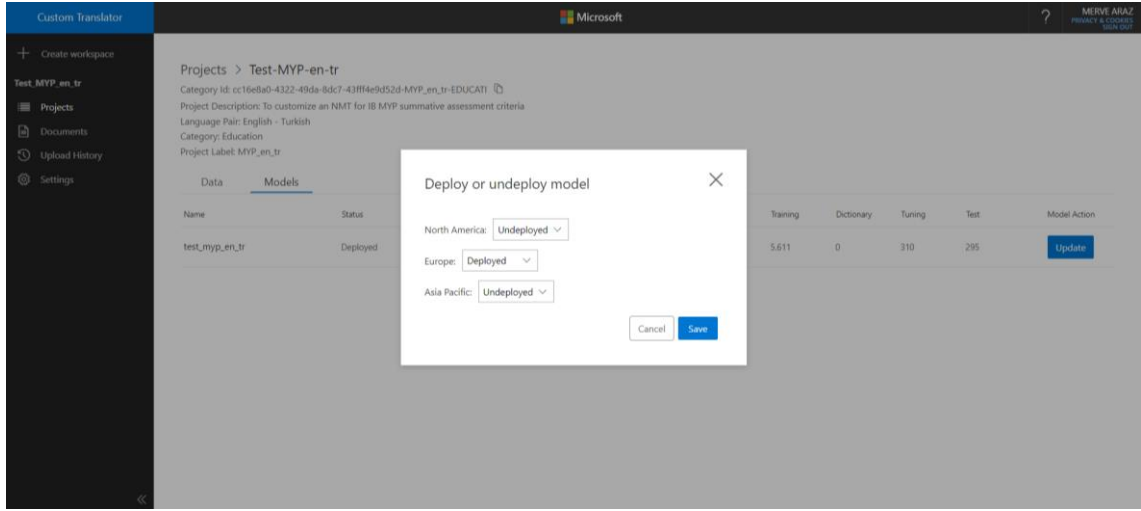


Figure 10: Starting the deployment of the model

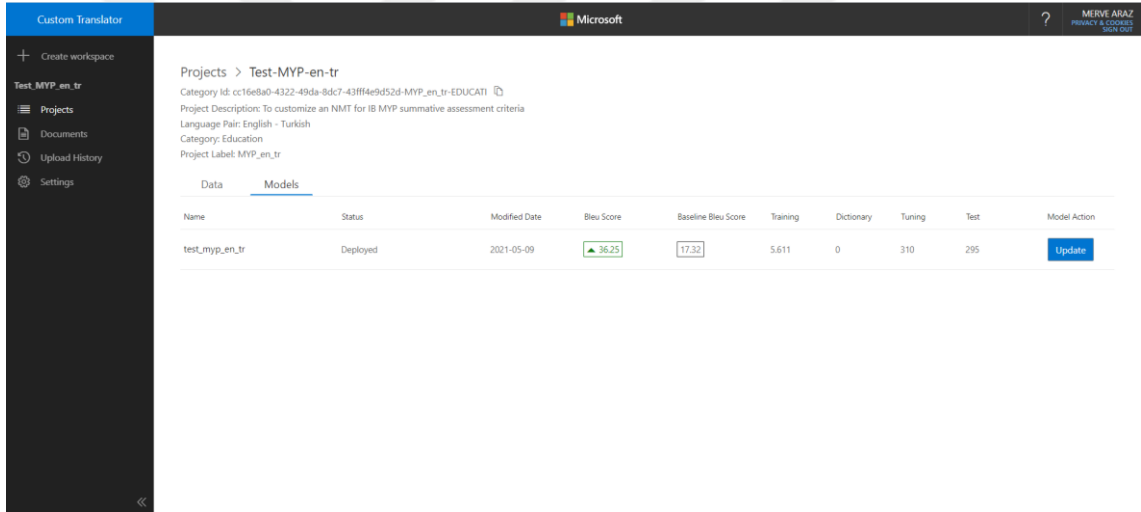


Figure 11: Deployment of the model

5.5 Using Deployed Model

Since this part of the study necessitates coding skills and software knowledge, a professional support is received from Tuğbanur Büte¹⁵ who is a graduate of the Department of Computer Education and Instructional Technologies at Yıldız Technical University. To this end, o collaboration is made and codes are provided by me while support is received from her computational expertise.

¹⁵ Tuğbanur Büte graduated from the Department of Computer Education and Instructional Technologies at Yıldız Technical University in 2021.

A call is required to make to Microsoft Translator Text API (Application Programming Interface) to be able to use the customized engine. By using the subscription key and determining the source and target languages, it is possible to obtain translations through the engine. To this end, two software are needed to be installed: Node.js and Visual Studio Code.

5.5.1 Node.js

Node.js is a JavaScript runtime built on Chrome's V8 JavaScript engine¹⁶. It is not a programming language but a Java Runtime Environment (JRE). JRE is a software that is made to execute other software¹⁷. It is available for Windows and macOS. Within this context, Node.js for Windows 64 is installed.

5.5.2 Visual Studio Code

In order to work on Node.js, an Integrated Development Environment (IDE) is required. An IDE enables developers to write and edit computer programmes. Visual Studio Code is a quite popular IDE which is lightweight and has a basic structure. It is available for macOS, Linux, and Windows. Visual Studio Code includes enriched built-in support for Node.js development with JavaScript¹⁸. Therefore, Visual Studio Code is installed.

5.5.3 Running Deployed Model

Upon the completion of installments of the software, a new file is opened in Visual Studio Code. Later, the following commands are entered.

```
const request = require('request');
const {v4} = require('uuid');
var key_var = 'SUBSCRIPTION_KEY';
var location = "global";
var endpoint = 'https://api.cognitive.microsofttranslator.com/';
function translateText() {
  let options = {
```

¹⁶ “Home”, Node.js, accessed May 23, 2021. <https://nodejs.org/en/>

¹⁷ “Is Node JS a Framework? Learn About the Popular Platform”, Effectus Software, accessed May 23, 2021. <https://effectussoftware.com/blog/node-js-a-framework/>

¹⁸ “Why Did We Build the Visual Studio Code?”, Visual Studio Code, accessed May 23, 2021. <https://code.visualstudio.com/docs/editor/whyvscode>

```

method: 'POST',
baseUrl: endpoint,
url: 'translate',
qs: {
  'api-version': '3.0',
  'category': 'MODEL_CATEGORY',
  'from': ['en'],
  'to': ['tr'],
},
headers: {
  'Ocp-Apim-Subscription-Key': key_var,
  'Content-type': 'application/json',
  'X-ClientTraceId': v4()
},
body: [{
  'text': 'SOURCE_TEXT'
}],
json: true,
};

request(options, function (err, res, body) {
  console.log(JSON.stringify(body, null, 4));
}); }; translateText()

```

In order to explain what has been done, above-given commands and their functions are explained below. Firstly, a request call is made to Microsoft Translator API by writing the first two commands:

```

const request = require('request');
const {v4} = require('uuid');

```

Later, the subscription key of the model is entered via the following command:

```

var key_var = 'SUBSCRIPTION KEY'

```


Next, the location of the model that is selected when creating the workspace and endpoint is commanded:

```
var location = "global";  
var endpoint = 'https://api.cognitive.microsofttranslator.com/';
```

Afterwards, the API version is entered:

```
'api-version': '3.0',
```

Then, the category model is commanded. In this model, the category is Education. Later, the source and target languages are written. Source is written with ‘from’ command and target with ‘to’ command.

```
'from': ['en'],  
'to': ['tr'],
```

Next to the ‘text’ command, the source text is written within the apostrophes as shown below:

```
'text': 'SOURCE_TEXT'
```

At the end of the above-explained process, the interface of the system appears as follows:

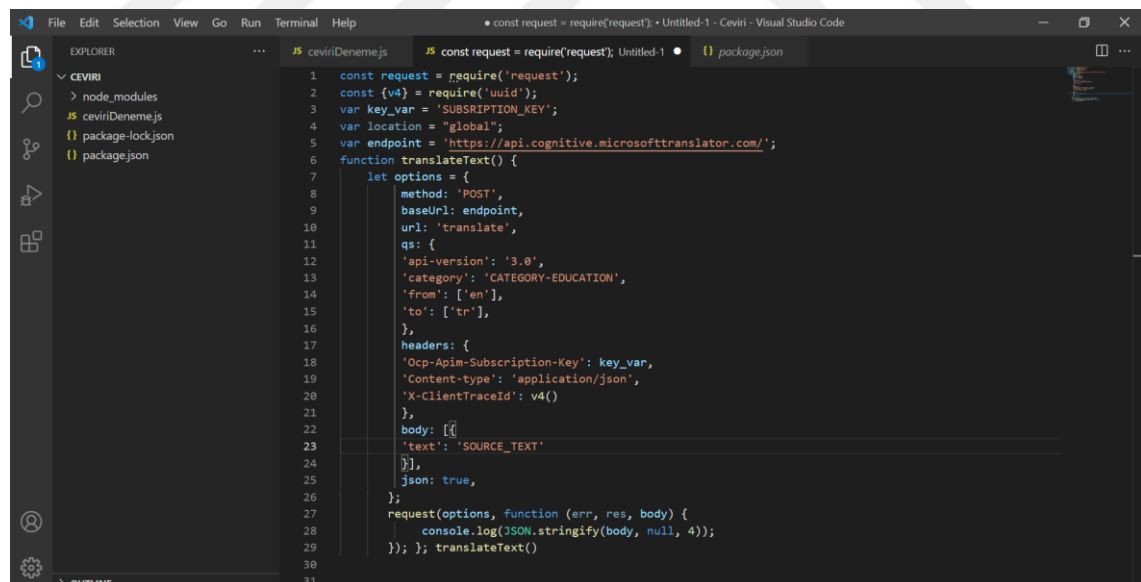
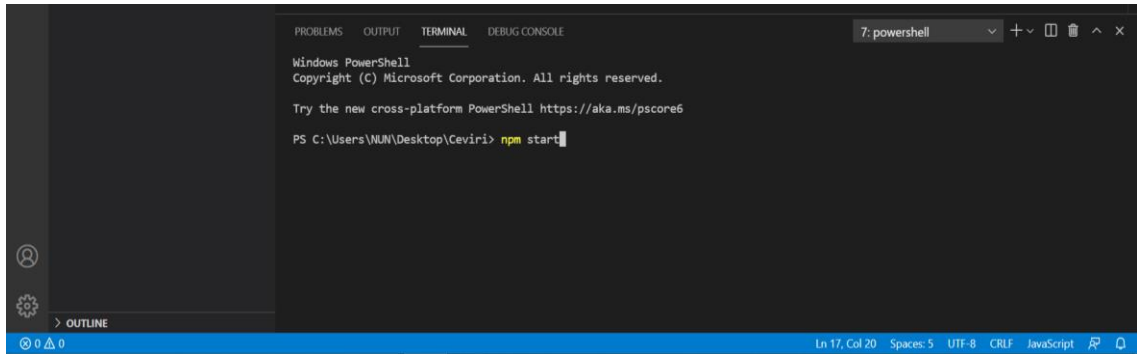


Figure 12: Visual Studio Code Interface

After these steps are taken, the system becomes viable to conduct translations. To this end, source text is entered in the “‘text’:” command. Then, “Terminal > New

Terminal” is selected from the ribbon. After the Terminal window opens, the “npm start” command is written and enter key is clicked:



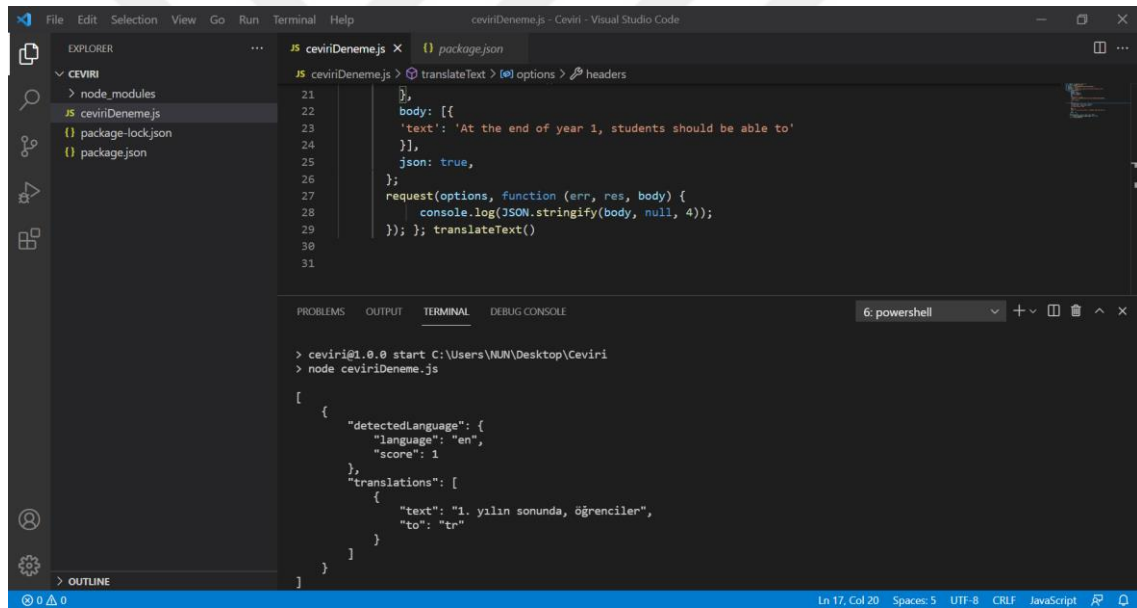
```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\NUN\Desktop\Ceviri> npm start
```

Figure 13: Visual Studio Code Translation Command

Finally, the target text appears on the Terminal screen as shown in the Figure 13.



```
> ceviri@1.0.0 start C:\Users\NUN\Desktop\Ceviri
> node ceviriDeneme.js

[
  {
    "detectedLanguage": {
      "language": "en",
      "score": 1
    },
    "translations": [
      {
        "text": "1. yılın sonunda, öğrenciler",
        "to": "tr"
      }
    ]
  }
]
```

Figure 14: Visual Studio Code Translation Output

The process of running and using Microsoft Custom Translator is explained meticulously in the above sections. Using these steps, MYP summative assessment materials can be translated via Microsoft Custom Translator. However, to assess the quality of the translations, a system or a framework is needed. Therefore, the Multidimensional Quality Metrics framework is used in this study to evaluate the quality of the acquired outputs. The following chapter focuses on the Multidimensional Quality Metrics.

6. MULTIDIMENSIONAL QUALITY METRICS

In order to determine whether an output is eligible to use, the quality of the translation needs to be ascertained. With this end, EU-funded QTLaunchPad project developed the Multidimensional Quality Metrics (MQM). The MQM is a framework that includes definitions and descriptions used for assessing the quality of translations. In this thesis, the Multidimensional Quality Metrics are going to be applied to assess the quality of machine translation outputs.

Since it contains a large number of issues that can be applied to measure the translation quality, the MQM defined a “‘Core’ consisting of 20 issue types that represent the most common issues arising in quality assessment of translated texts”¹⁹. The graphical representation of the MQM Core is as follows:

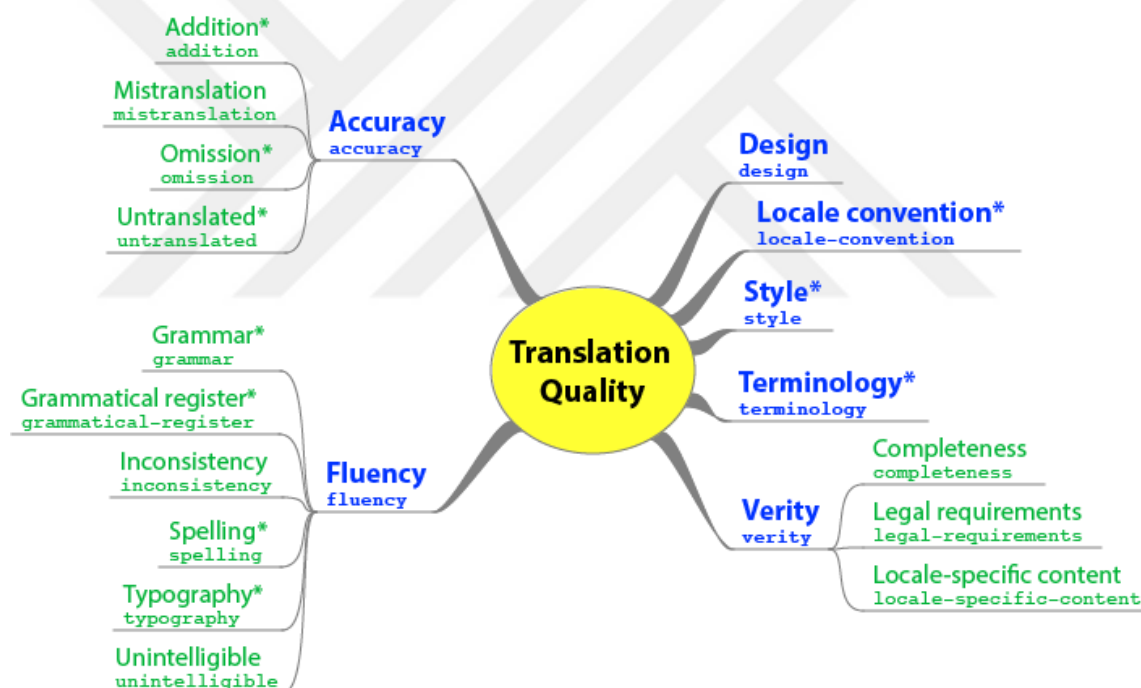


Figure 15: MQM Core Representation

Source: <https://www.qt21.eu/mqm-definition/definition-2015-12-30.html#scoring-algorithm>

¹⁹ Multidimensional Quality Metrics Definition, accessed August 29, 2021. <https://www.qt21.eu/mqm-definition/definition-2015-12-30.html#scoring-algorithm>

Each of these issues can be applied in accordance with the needs of the translation document. In order to further elaborate on the issues listed in the MQM Core, their extensive definitions are provided below.

a) Accuracy:

Accuracy addresses to the issues that deal with the relationship of the target text and the source text. Accuracy embraces a list of issues that are addition, mistranslation, omission, over-translation, under-translation and untranslated (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

b) Fluency:

Fluency deals with issues about the language of the translation. It may be applied to both source and target texts. Fluency issues can be listed as follows: ambiguity, character encoding, coherence, cohesion, corpus conformance, duplication, grammar, grammatical register, inconsistency, index/table of contents format, link/cross reference, non-allowed characters, offensive, pattern problem, sorting, spelling, typography, unintelligible (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

c) Design:

Design tackles with the physical features of the target text. In other words, it concentrates on matters about the layout of the translation. The list of the design issues encapsulates graphics and tables, hyphenation, length, local formatting, markup, missing text, truncation/text expansion, overall design (layout) (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

d) Locale convention:

Locale convention issues are related to the compliance of the content with the “locale-specific conventions”. If the source text is correctly rendered in the target text and the target text is fluent but has minor locale-specific mistakes, it is examined under this topic. Locale convention comprises the following issues: address format, calendar type, currency format, date format, locale-specific punctuation, measurement format, name format, national language standard, number format, shortcut keys, telephone format, time

format (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

e) Style:

Style issues are related to stylistic features of the target text. It is also related to the fluency. Style issues can be listed as register, third-party style, unidiomatic, inconsistent style, company style, awkward (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

f) Terminology:

Terminology issues deal with domain specific terminology usage. Any terminological problem must be included in this section. The list of terminology issues is as follows: inconsistent with termbase, inconsistent with domain, inconsistent use of terminology (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

g) Verity:

Verity issues are linked with suitability of the target text with the audience and context. The accuracy of a text does not ensure its verity, too. Verity issues can be listed as completeness, culture-specific reference, end-user suitability, legal requirements, locale-specific content (see <https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (accessed on September 5, 2021)).

The following chapter includes some examples of translations that are obtained from the NMT engine. These translations will be evaluated in light of the Core issues of the Multidimensional Quality Metrics.

7. CUSTOM TRANSLATOR OUTPUT ANALYSIS

Up to the present, training materials are collected to train the NMT engine, the Custom Translator is prepared to translate and the quality metrics are determined. At this stage, test materials that are going to be translated need to be specified.

When training the engine, the summative assessment criteria received from the IB MYP education model are used. Hence, actual summative assessments are chosen as test materials. These summative assessment criteria are assigned to students in a classroom. They are obtained from an IB Continuum School based in Istanbul, Turkey. Each document has been specifically prepared for the actual needs of the classroom that it is assigned to.

In total, 20 examples are acquired for the purposes of this thesis. Half of these examples are put to test to check whether the engine successfully works or not. The rest of the examples are examined within the scope of the following chapter. The passages obtained from summative assessments which are used in the previously mentioned IB Continuum School are provided under the Source Text column in the below given tables. The outputs obtained from the Custom Translator engine are shared under the Target Text column of the table. Beneath each example, an analysis of the translations is provided.

Example 1:

Source Text	Target Text
The student is able to: i. state a research question on the osmosis in eggs in different type of solutions to be tested by the experiment, with limited success ii. state a testable hypothesis of the osmosis in eggs experiment in different type of solutions iii. state the variables in the egg osmosis experiment between different type of solutions iv. design an experimental method of the osmosis in egg between different type of solutions, with limited success. (Science – Criterion B – Year 5)	Öğrenci şunları yapabiliyor: i. deney tarafından test edilecek farklı çözüm türlerinde yumurtalardaki ozmoz hakkında bir araştırma sorusu belirtmek, sınırlı başarı ile ii. yumurta deneylerinde ozmozun test edilebilir bir hipotezini farklı çözüm türlerinde belirtmek iii. yumurta ozmos deneyindeki değişkenleri farklı çözüm türleri arasında belirtmek iv. yumurtadaki ozmozun farklı çözüm türleri arasında deneysel bir yöntemini tasarlamak, sınırlı başarı ile

Comment:

In this example, it is seen that all four of the numbered sentences are translated consistently in infinitive form. Additionally, the word “osmosis” is translated consistently in terms of domain. This translation is rendered as correct in accordance with the terminology issue of MQM Core list. Nevertheless, the translation of the phrase “with limited success” is translated at the end of the sentences numbered i. and iv. In regards to syntax rules, this phrase is an adverb. Therefore, it would have been better if it had been placed right before the verb. In that case, the translation of the sentence numbered iv. would be as follows: *yumurtadaki ozmozun farklı çözüm türleri arasında deneysel bir yöntemi **sınırlı bir başarı ile** tasarlamak*. Even so, it does not lead to any inaccuracy or misunderstanding in the overall structure of the target text.

Example 2:

Source Text	Target Text
The student is able to: i. summarize the ways in which graphene as a material is applied and used to address the applications of graphene. ii. describe the implications of utilizing the graphene as a material and its applications to development of the countries. iii. sometimes apply scientific language to communicate understanding graphene and its properties. iv. sometimes document sources about graphene and its properties correctly. (in MLA format) (Science – Criterion D – Year 5)	Öğrenci şunları yapabiliyor: i. grafenin bir malzeme olarak uygulanma ve grafen uygulamalarını ele almak için kullanılma yollarını özetleyebiliyor. ii. grafenin bir malzeme olarak kullanılmasının etkilerini ve ülkelerin kalkınmasına yönelik uygulamalarını açıklamak. iii. bazen grafen ve özelliklerini anlamak için bilimsel dil uygular. iv. bazen grafen ve özellikleri hakkındaki kaynakları doğru bir şekilde belgelemektedir. (MLA biçiminde)

Comment:

In this example, source text information is accurately and completely conveyed to the source language. Area-specific terms are also correctly translated. However, verbs are not translated in a consistent manner. Verbs in the target text are translated in the following structures, respectively: present continuous tense, infinitive, present tense, present tense.

The inconsistency in the first two sentences reduces the fluency of the target text. However, this does not have any effect on the accuracy of the translation. With a simple post-editing process, this inconsistency can easily be eliminated.

Example 3:

Source Text	Target Text
The student is able to: i. select appropriate mathematics when solving simple problems in familiar situations ii. apply the selected mathematics successfully when identifying amount of each nutrition and its cost iii. generally identify amount of each nutrition and its cost correctly in a variety of contexts. (Mathematics – Criterion A – Year 5)	Öğrenci şunları yapabiliyor: i. tanıdık durumlarda basit sorunları çözerken uygun matematiği seçebiliyor ii. her beslenme miktarını ve maliyetini belirlerken seçilen matematiği başarıyla uygulayabiliyor iii. genellikle her beslenme miktarını ve maliyetini çeşitli bağlamlarda doğru bir şekilde tanımlayabiliyor.

Comment:

When Example 3 is reviewed in line with the MQM Core issues, it can be observed that the general flow of the target text is accurate and fluent regarding coherence and consistence. Verbs of the sentences are translated in the same tense: present tense. Besides, sentences that are given in numbered list comply with the syntax of the main sentence. However, the NMT engine mistranslated the word “nutrition” as *beslenme* which should have been translated as *besin* in this context. This should be considered within the scope of the accuracy issue.

Example 4:

Source Text	Target Text
The student is able to: i. outline scientific knowledge about environmental problems ii. apply scientific knowledge and understanding to solve environmental problems set in familiar situations iii. interpret information to make scientifically supported judgments about environmental problems.	Öğrenci şunları yapabiliyor: i. çevre sorunları hakkında bilimsel bilgileri ana hatlarıyla belirlemek ii. bilindik durumlarda belirlenen çevresel sorunları çözmek için bilimsel bilgi ve anlayışı uygulamak iii. çevre sorunları hakkında bilimsel olarak desteklenen yargılarda

(Science – Criterion A – Year 5)	bulunmak için bilgileri yorumlamak
----------------------------------	------------------------------------

Comment:

As this example is analyzed with a view to MQM issues, it can be stated that the source text is transferred to the target language in an accurate and fluent manner. All of the sentences are consistent in regards to syntactic structure. Yet, there is a phrase that is translated word-for-word in the sentence numbered iii. This expression is “scientifically supported judgements” and it is translated as *bilimsel olarak desteklenen yargılar*. This translation of the phrase may disrupt the overall intelligibility of the sentence. Instead, translating it as *bilimsel yargılar* could have had a more positive affect on the comprehensibility of the statement.

Example 5:

Source Text	Target Text
The student is able to:	Öğrenci şunları yapıyor:
i. identify the relevant elements of test results of cancer	i. kanserin test sonuçlarının ilgili unsurlarını belirlemek
ii. select adequate mathematical strategies to identify whether a patient has cancer or not	ii. bir hastanın kanser olup olmadığını belirlemek için yeterli matematiksel stratejileri seçmek
iii. apply the selected mathematical strategies to identify whether a patient has cancer or not, by reaching a correct solution	iii. doğru bir çözüme ulaşarak, bir hastanın kanser olup olmadığını belirlemek için seçilen matematiksel stratejileri uygulamak
iv. explain the degree of accuracy of the solution of identifying whether a patient has cancer or not	iv. bir hastanın kanser olup olmadığını belirleme çözümünün doğruluk derecesini açıklamak
v. describe correctly whether the solution makes sense in the context of the authentic real-life situation	v. çözümün gerçek yaşam durumu bağlamında anlamlı olup olmadığını doğru bir şekilde tanımlamak
(Mathematics – Criterion D – Year 5)	

Comment:

In Example 5, target text is in compliance with the accuracy and fluency issues in the framework of the MQM. Translation is coherent and understandable. When examined

from the point of syntax, it can be claimed that all of the numbered sentences are translated in infinitive-verb form. This ensures and enhances the consistency of the target text. A similar situation observed in the Example 3, sentences numbered i. and iv., adverbs are correctly placed in this example. For instance, in the target sentence numbered v., it is seen that the adverb *doğru bir şekilde* comes before the verb *tanımlamak*. This structure is grammatically correct in Turkish.

Example 6:

Source Text	Target Text
The student is able to: i. state the ways in which science is used to address a specific problem or issue on vitamins and minerals. ii. state the implications of using science to solve a specific problem or issue, interacting with vitamins and minerals. (Science – Criterion D – Year 5)	Öğrenci şunları yapıyor: i. bilimin vitamin ve minerallerle ilgili belirli bir sorunu veya sorunu ele almak için nasıl kullanıldığını belirtmek ii. belirli bir sorunu veya sorunu çözmek için bilimi kullanmanın, vitamin ve minerallerle etkileşimde bulunmanın etkilerini belirtmek.

Comment:

When looked at this example, a consistency can be observed in the syntactic structure of the target text. Text is fluent, intelligible and accurate. However, there are minor i.e. word-level translation problems. For instance, both of the words “problem” and “issue” are translated into Turkish as *sorun*. Since these words are used in the source text as “problem or issue”, they are transferred as *sorun ve sorun* which interrupts the fluency of the target text. Instead, Turkish words such as *konu*, *mesele*, *problem* could have been preferred for the translation of these near-synonymous words in order to prevent using synonyms successively.

Example 7:

Source Text	Target Text
The student is able to: i. usually use appropriate mathematical and academic language while explaining mathematical and logical steps to identify similarity relations.	Öğrenci şunları yapıyor: i. benzerlik ilişkilerini tanımlamak için matematiksel ve mantıksal adımları açıklarken genellikle uygun matematiksel ve akademik dili kullanmak

ii. usually use appropriate forms and symbols of mathematical representation to identify similarity relations.	ii. benzerlik ilişkilerini tanımlamak için genellikle matematiksel temsilin uygun formlarını ve sembollerini kullanmak
iii. communicate through lines of reasoning that are able to be understood for identifying similar triangles.	iii. benzer üçgenleri tanımlamak için anlaşılabilen akıl yürütme çizgileri aracılığıyla iletişim kurmak
iv. present work that is usually organized using a logical structure to identify similar triangles in the example.	iv. genellikle örnekteki benzer üçgenleri tanımlamak için mantıksal bir yapı kullanılarak düzenlenmiş çalışmaları sunmak
(Mathematics – Criterion C – Year 1)	

Comment:

In this example, target text is accurate as the source text information is transferred to the target language precisely. The syntactic structure of the target text is consistent within itself. It is easily noted that the translation is quite intelligible. However, there is a minor translation issue in the target text. In the sentence numbered iii., the source text expression “lines of reasoning” is translated into Turkish as *akıl yürütme çizgileri* in a word-for-word translation approach. This translation disrupts the fluency of the target text. To use *akıl yürütmek* or *akıl yürüterek* would have sound more natural in Turkish. Although this does not change or affect the meaning, this type of an editing may increase the fluency of the text.

Example 8:

Source Text	Target Text
The student is able to, with limited success:	Öğrenci şunları yapıyor:
i. state the ways in which science is used to address a health problem or issue on the organelles losing their functions, mentioning advantages and disadvantages of organelles presence.	i. organel varlığının avantaj ve dezavantajlarından bahsederek, organellerin işlevlerini kaybetmesiyle ilgili bir sağlık sorununu veya sorununu ele almak için bilimin nasıl kullanıldığını belirtmek,
ii. state the implications of using science and its application to solve a health problem or issue, interacting with organelles.	ii. bir sağlık sorununu veya sorununu çözmek için bilimi ve uygulamasını kullanmanın etkilerini belirtmek

iii. apply scientific language to communicate with class community about your cell module.	iii. organellerle etkileşime girmek, hücre modülünüz hakkında sınıf topluluğuyla iletişim kurmak için bilimsel dil uygulamak
iv. document sources about cell structure, organelles and their functions. (Science – Criterion D – Year 5)	iv. hücre yapısı, organeller ve işlevleri hakkında kaynakları belgeleme.

Comment:

In accordance with the issues defined in the MQM, the translation given in the Example 8 is accurate and fluent. Domain specific terms are correctly and consistently translated in regard to terminology issue of the given quality metrics. As in the Example 6, there is a mistake in the translation of near-synonymous words “problem” and “issue”. Since both of them are translated as *sorun*, the phrase turns out to be *sorun ve sorun* which needs to be post-edited before the text is used. The Custom Translator maintained the consistency in the translation of the verbs by using infinitive-verb structure. There are three types of infinitives in Turkish: *-ma (-me)*, *-mak, (-mek)*, *-ış, (-ış, -uş, -üş)*. In the first three sentences, “*-mak, (-mek)*” structure is preferred. However, the infinitive form *-ma (-me)* is preferred in the last sentence. This situation slightly interrupts the fluency of the text in spite of the fact that it does not have an effect on the meaning of the text.

Example 9:

Source Text	Target Text
The student is able to:	Öğrenci şunları yapıyor:
i. select appropriate mathematics when solving challenging problems in familiar situations	i. tanıdık durumlarda zorlu sorunları çözerken uygun matematiği seçmek
ii. apply the selected mathematics successfully when identifying the most appropriate design	ii. en uygun tasarımı belirlerken seçilen matematiği başarıyla uygulamak
iii. generally identify the number of cars in parallel, perpendicular, and angled parking (Mathematics – Criterion A – Year 5)	iii. genellikle paralel, dik ve açılı park halindeki araç sayısını belirlemek

Comment:

In this example, the translation is in accordance with the MQM issues. Target text is accurate, fluent and the terms are transferred correctly in respect to the domain. In the source text sentence numbered iii., the adverb “generally” defines the verb “identify”. However, in the target text sentence numbered iii., the expression *genellikle* which is the equivalent of “generally” does not directly qualify the verb *belirlemek* but the adjectival phrase *paralel, dik ve açılı park halindeki araç sayısı*. In order to prevent this situation, the phrase *genellikle* can be placed before the verb *belirlemek*. In such a case, the target sentence would be *paralel, dik ve açılı park halindeki araç sayısını genellikle belirlemek*.

Example 10:

Source Text	Target Text
The student is able to: i. state scientific knowledge on Newton’s Laws and motion ii. apply scientific knowledge and understanding to suggest solutions to problems set in familiar situations related to forces and motion. iii. interpret information to make judgments in questions 2,3,4 and 5. (Science – Criterion A – Year 5)	Öğrenci şunları yapıyor: i. Newton Yasaları ve önergesi hakkında bilimsel bilgi vermek ii. kuvvetler ve hareketle ilgili tanıdık durumlarda belirlenen sorunlara çözüm önermek için bilimsel bilgi ve anlayışı uygulamak iii. 2,3,4 ve 5. sorularda yargılama yapmak için bilgileri yorumlamak.

Comment:

In this example, there are a couple of issues that need to be examined. Firstly, the word “motion” in the source text sentence numbered i. is mistranslated as *önerge*. Although *önerge* is one of the equivalents of the word motion, in this context, it amounts to movement. Therefore, it should have been translated as *hareket*. This mistake can be evaluated under the accuracy > mistranslation and terminology > inconsistent with domain issues. Interestingly, the same word is repeated in the following sentence and in this case it is correctly translated as *hareket*. The second issue in this example is detected in the sentence numbered iii. The expression “make judgments” is translated as *yargılama*

yapmak. In this context, this statement means to decide. Hence, it could have been conveyed in this direction. This type of a solution could have enhanced the fluency of the target sentence.

7.1 Custom Translator Output Analysis Evaluation

In the previous section, various source texts are put to test the Custom Translator. Then, outputs obtained from the NMT engine are shared together with the source texts. Each translation is analyzed in detail in accordance with MQM Core issues list. In these analyses, three MQM issues step forth: accuracy, fluency, terminology.

First and foremost, it is observed that all of the source texts are translated accurately. Although a couple of slight mistakes are detected, there is no target text that contains misinformation. Secondly, consistency is maintained in almost all translations with regard to syntactic and grammatical structure. This situation further extended the fluency in the target texts. Thirdly, it is found out that the Custom Translator is quite good at terminology translation. Only one term was mistranslated, yet it was correctly translated in another sentence.

However, some inconsistencies are found in terms of terminology. In addition, some hurdles such as precarious use of tenses and infinitive verbs prevented better fluency rate in the target text. Therefore, Custom Translator needs to be improved to produce more fluent texts. To this end, more training data could be uploaded to the Custom Translator. Apart from this analysis, the quality of the translation is analyzed through a different aspect: BLEU Score metrics. More detailed information on BLEU Score can be found in the following section.

7.2 BLEU Score of the Customized NMT Outputs

In order to evaluate the quality of the machine translation output, an automatic evaluation system (i.e. automated metric system) needs to be used. Automated metric systems emerged as an alternative to human evaluation which is time-consuming and expensive. Several automated metric systems are available currently. In this study, Bilingual Evaluation Understudy (BLEU), proposed by Papineni and et al. in 2002, metric system is used. BLEU score mainly focuses on two factors when evaluating machine translation output: the closeness of output to the reference translation and the quality of reference translation (see Papineni et al. 2002, 311). Reference translation is the human-produced

good quality translation. The closer the output to the reference translation, the higher is the BLEU score. BLEU metric ranges from 0 to 1 (Papineni et al. 2002, 315). The number of reference translations affect the BLEU score. In other words, the more reference translations there are, the higher is the BLEU score. The quality of the reference translation directly has an impact on the BLEU score.

Microsoft Custom Translator has a built-in BLEU Score metric system. According to the result obtained from that built-in BLUE Score, the baseline BLEU Score of the customized NMT engine is 17.32 while the BLEU Score of the custom engine is 36.25 which is interpreted as follows in accordance with the BLUE metric scale:

Table 1
BLEU Score Interpretation

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

Source: <https://cloud.google.com/translate/automl/docs/evaluate>

In light of the information given on the table, the BLEU Score of the custom model is “understandable to good translations”. Although BLEU Score is a proof of the quality of the engine, in the following chapter several additional examples will be shared to further examine the quality of the outputs created by the customized engine. In so doing, SMT outputs will also be taken into consideration.

8. COMPARISON OF SMT AND NMT OUTPUTS

The previous chapter addresses the translation performance of the Custom Translator. In order to further develop the analysis towards translation quality of the NMT output, a number of source texts are put to translation via a customized neural machine translation engine and two different generic SMT systems. To this end, for the customized engine, the Microsoft Custom Translator is used. It is the only neural machine translation engine included into the comparison. Google also has a separate NMT system which is called Google AutoML Translator. However, it supports a limited number of file formats. Hence, it could not be able to be trained with the same amount of documents that are used to train the Microsoft Custom Translator. Since comparing two different NMT that are trained with unequal amount of data would not be fair, Google AutoML is not included in this comparison.

As to the SMT engines, Google Translate, Bing Microsoft Translator, Yandex Translate, iTranslate, Reverso Translate, etc. some of the well-known systems are involved in the comparison process. Since Custom Translator is a product of Microsoft, Bing Microsoft Translator is not selected not to use the two products of the same company. iTranslate, Reverso Translate and other similar translation systems are not widespread as much as Google Translate and Yandex Translate. Therefore, they are not selected as well. In the end, it is decided to choose Google and Yandex's SMT systems for the comparison process. All the systems of which outputs are to be compared are labelled as follows: Microsoft Custom Translator (NMT), Google Translate (SMT-1) and Yandex Translate (SMT-2).

Below are the source text examples obtained from test materials that are acquired from an IB Continuum School based in Istanbul, Turkey. These materials are prepared in an actual teaching environment and assigned to the students. In other words, authentic texts are used to test the machine translation systems.

Example 11:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. collect and present data in the	Öğrenci şunları yapabiliyor: i. sayısal ve görsel formlarda daha	Öğrenci şunları yapabilir: ben. Sayısal ve görsel formlarda	Öğrenci şunları yapabilir:. i. sayısal ve görsel formlarda daha

experiment between more viscous and less viscous solutions in numerical and visual forms ii. accurately interpret data in the reaction between more viscous and less viscous solutions iii. state the validity of the hypothesis of the reaction between more viscous and less viscous solutions with limited reference to the experiment iv. state the validity of the method used in the reaction between more viscous and less viscous solutions with limited reference to the experiment (<i>Science – Criterion C – Year 5</i>)	viskoz ve daha az viskoz çözümler arasındaki deneyde veri toplamak ve sunmak ii. daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyondaki verileri doğru bir şekilde yorumlamak iii. deneye sınırlı referansla daha viskoz ve daha az viskoz çözümler arasındaki reaksiyon hipotezinin geçerliliğini belirtmek iv. deneye sınırlı referansla daha viskoz ve daha az viskoz çözümler arasındaki reaksiyonda kullanılan yöntemin geçerliliğini belirtmek	daha viskoz ve daha az viskoz çözümler arasındaki deneyde veri toplamak ve sunmak ii. Daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyondaki verileri doğru şekilde yorumlayın iii. Deneye sınırlı referansla daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyon hipotezinin geçerliliğini belirtin iv. Deneye sınırlı referansla daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyonda kullanılan yöntemin geçerliliğini belirtin	viskoz ve daha az viskoz çözeltiler arasındaki deneyde veri toplamak ve sunmak ii. daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyondaki verileri doğru bir şekilde yorumlar iii. deneye sınırlı referansla daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyon hipotezinin geçerliliğini belirtin iv. deneye sınırlı referansla daha viskoz ve daha az viskoz çözeltiler arasındaki reaksiyonda kullanılan yöntemin geçerliliğini belirtir
---	--	---	--

Comment:

To begin with the translation of the first sentence, NMT conveyed it in the present continuous tense while SMT-1 and SMT-2 referred to the present tense. Unlike the other outputs, there are two colons in the first sentence of the target text obtained through SMT-2. It can be included in the punctuation subheading under the locale-convention issue of the MQM.

In the SMT-2, the number “i.” is conveyed as *ben* in the target text. It is not translated as a number but as the translation of the pronoun “I”. This mistake also falls

under the locale convention issue of the MQM. A similar problem is observed in the SMT-2. The number of the fourth target sentence is not conveyed correctly. It is translated as “iv.” This transfer may also contradict with the number format of the locale-convention.

The word “solutions” in the source text sentence numbered ii. is translated as *çözüm*, *çözümlü*, *çözümlü* in NMT, SMT-1, SMT-2, respectively. In this context, the correct translation is *çözümlü*. In this case, the translation of this word provided by the NMT is inconsistent with domain and this issue is addressed under the terminology issue of the MQM. However, the same word is seen in the source sentence numbered iii. and it is translated as *çözümlü* in all target texts.

When looked at the general structure of the target texts, it is noticed that the output of the NMT has a consistent grammatical structure by using infinitive-verbs. In the SMT-1, sentences are translated in given structures as follows respectively: infinitive, imperative, imperative, imperative. Although there is a consistency in the translations of the verbs of the last three sentences, due to the grammatical structure of the very first sentence, it could be more appropriate to render those in the infinitive form. In the SMT-2, the translations of these verbs are done as follows: infinitive, present tense, imperative, present tense. The inconsistency is greater in this target text and it needs to be edited to have a more fluent text. However, these mistranslations do not change the meaning of the text and they can be corrected with a slight post-editing process.

To sum up, all the MT systems put forth accurate translations successfully. In terms of content, they also created fluent texts. Yet, when looked from the grammatical perspective, only NMT produced a fluent text. In addition, NMT output has a mistake that needs to be addressed under terminology issue while SMT-1 and SMT-2 outputs have some issues that needs to be examined under the locale-convention issue of the MQM.

Example 12:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. select and apply mathematical and logical	Öğrenci şunları yapabiliyor: i. Doğru kalıpları tanıyarak her beslenme miktarını ve maliyetini	Öğrenci şunları yapabilir: ben. Doğru kalıpları tanıyarak her beslenmenin miktarını ve	Öğrenci şunları yapabilir: i. doğru kalıpları tanıyarak her beslenmenin miktarını ve

steps to identify amount of each nutrition and its cost by recognizing correct patterns ii. describe a general rule consistent with identifying amount of each nutrition and its cost iii. verify whether these mathematical and logical thinking steps to identify amount of each nutrition and its cost work for other mixtures such as milkshake and soup (Mathematics – Criterion B- Year 5)	belirlemek için matematiksel ve mantıksal adımlar seçmek ve uygulamak ii. her beslenme miktarını ve maliyetini belirlemekle tutarlı genel bir kural tanımlamak iii. her beslenmenin miktarını ve maliyetini belirlemek için bu matematiksel ve mantıksal düşünme adımlarının milkshake ve çorba gibi diğer karışımlar için işe yarayıp yaramadığını doğrulamak	maliyetini belirlemek için matematiksel ve mantıksal adımları seçin ve uygulayın ii. Her bir besinin miktarını ve maliyetini belirleyen genel bir kural tanımlayın iii. Her bir besinin miktarını ve maliyetini belirlemek için bu matematiksel ve mantıksal düşünme adımlarının milkshake ve çorba gibi diğer karışımlar için işe yarayıp yaramadığını doğrulayın.	maliyetini belirlemek için matematiksel ve mantıksal adımları seçin ve uygulayın ii. her beslenmenin miktarını ve maliyetini belirlemeye uygun genel bir kural tanımlayın iii. her beslenmenin miktarını ve maliyetini belirlemek için bu matematiksel ve mantıksal düşünme adımlarının milkshake ve çorba gibi diğer karışımlar için işe yarayıp yaramadığını doğrulayın
---	--	---	---

Comment:

In this example, the same mistranslation of number that is seen in the previous example is found again in SMT-1. The number “i.” is translated as first person singular pronoun *ben* in Turkish. This is an incoherence with the accuracy issue of the MQM.

Furthermore, it can be claimed that the translation of the word “nutrition” created a confusion in this example. In this context, nutrition is an equivalent of food, nutriment, etc. Its Turkish counterpart is *besin*. However, it is translated as *beslenme* instead of *besin* in NMT and SMT-2 outputs. In SMT-1 target text, both *besin* and *beslenme* words are used to meet this meaning. These also can be classified under the mistranslation subheading of accuracy issue of the MQM.

As the grammatical structure of the target texts are examined, it is noted that all three of them are consistent. In NMT output, infinitive is used in the verb translation while imperative is applied in SMT-1 and SMT-2. However, infinitive usage is more

appropriate in this context rather than imperatives. Since it does not have any significant effect on meaning, it can be used or post-edited, if required.

In brief, all three have slight accuracy issues due to the expression “nutrition”. Yet, it can be stated that target texts are fluent and consistent in general.

Example 13:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. describe the problem you stated in part (a). You should include background information and cite your resources. ii. formulate and explain a testable hypothesis using scientific reasoning. You should state a relationship between variables. You may include your findings from your research and cite the resources. iii. describe how to manipulate the variables. iv. design a complete and safe method to reach the aim (Science – Criterion B – Year 5)	Öğrenci şunları yapabiliyor: i. belirttiğiniz sorunu kısmen (a) açıklar. Arka plan bilgilerinizi eklemeli ve kaynaklarınızı belirtmelisiniz. ii. Bilimsel akıl yürütmeyi kullanarak test edilebilir bir hipotez oluşturmak ve açıklamak. Değişkenler arasında bir ilişki belirtmelisiniz. Araştırmanızdan elde ettiğiniz bulguları dahil edebilir ve kaynakları alıntılatabilirsiniz. iii. değişkenlerin nasıl manipüle edilebileceğini açıklayabilir iv. amaca ulaşmak için eksiksiz ve güvenli bir yöntem tasarlayabilirsiniz	Öğrenci şunları yapabilir: ben. (a) kısmında belirttiğiniz sorunu tanımlayın. Arka plan bilgilerini eklemeli ve kaynaklarınızdan alıntı yapmalısınız. ii. Bilimsel akıl yürütmeyi kullanarak test edilebilir bir hipotez formüle edin ve açıklayın. Değişkenler arasında bir ilişki belirtmelisiniz. Araştırmanızdan elde ettiğiniz bulguları ekleyebilir ve kaynakları belirtebilirsiniz. iii. değişkenlerin nasıl manipüle edileceğini açıklayın. iv. amaca ulaşmak için eksiksiz ve güvenli bir yöntem tasarlamak	Öğrenci şunları yapabilir:: i.bölüm (a) ' da belirttiğiniz sorunu açıklayınız. Arka plan bilgilerini eklemeli ve kaynaklarınızı belirtmelisiniz. ıı. bilimsel akıl yürütmeyi kullanarak test edilebilir bir hipotezi formüle eder ve açıklar. Değişkenler arasında bir ilişki belirtmelisiniz. Bulgularınızı araştırmanıza dahil edebilir ve kaynakları aktarabilirsiniz. ııı. değişkenlerin nasıl manipüle edileceğini açıklayın. iv. hedefe ulaşmak için eksi

Comment:

In this example, the expression “part (a)” mistranslated in NMT while it is accurately translated in SMT-1 and SMT-2. The additional statements given in sentences numbered i. and ii. are correctly conveyed in all of the machine translation outputs. SMT-1 did not translate “i.” correctly. The number is translated as if it is a pronoun in Turkish. This problem is observed only in SMT-1.

The phrase “formulating ... hypothesis” is given as *hipotez oluşturmak* in NMT in a manner that suits the nature of Turkish. But, it is translated as *hipotez formüle etmek* in SMT-1 and SMT-2. It is an example of word-for-word translation and it does not sound as a natural expression in Turkish. However, the message can still be obtained through this statement. Hence, it may not be a major issue.

In SMT-2, the sentence numbered iv. is not translated. This matter is evaluated under the untranslated heading of accuracy and it is a serious mistake that should be taken into consideration.

Example 14:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. identify the relevant elements of city planning (<i>Mathematics - Year 5 – Criterion D</i>)	Öğrenci şunları yapabilmektedir: i. şehir planlamasının ilgili unsurlarını belirlemek	Öğrenci şunları yapabilir: ben. şehir planlamasının ilgili unsurlarını belirlemek	Öğrenci şunları yapabilir: i. şehir planlamasının ilgili unsurlarını tanımlayın

Comment:

In the above example, NMT and SMT-1 produced similar target texts. The only mistake of the SMT-1 is in the translation of numbers. SMT-1 translated the number as a Turkish pronoun: *ben*.

When it comes to the SMT-2, it can be realized that there is a conflict between the translations of the verbs of the two sentences. While former verb is correctly translated

in the modal verb form, the second one translated in the imperative form. This situation disrupts the fluency of the text and it is a matter that should be addressed under the fluency issue of the MQM.

Example 15:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. collect and present data in the experiment between different metals and metal's solutions in numerical and visual forms ii. accurately interpret data in the reaction between different metals and metal's solutions (<i>Science Year 3 – Criterion C</i>)	Öğrenci şunları yapabiliyor: i. farklı metaller ve metal çözümleri arasındaki deneyde sayısal ve görsel formlarda veri toplayıp sunabiliyor ii. farklı metaller ve metal çözümleri arasındaki reaksiyondaki verileri doğru bir şekilde yorumlamak	Öğrenci şunları yapabilir: ben. Farklı metaller ve metal çözümleri arasındaki deneyde sayısal ve görsel formlarda veri toplamak ve sunmak ii. Farklı metaller ve metal çözümleri arasındaki reaksiyondaki verileri doğru şekilde yorumlayın	Öğrenci şunları yapabilir: i. sayısal ve görsel formlarda farklı metaller ve metal çözümleri arasındaki deneyde veri toplamak ve sunmak ii. farklı metaller ve metal çözümleri arasındaki reaksiyondaki verileri doğru bir şekilde yorumlayın

Comment:

In the above example, SMT-1 mistranslated the number “i.” into Turkish. It is generated as the first-person singular pronoun. Albeit, all three engines successfully translated the verb of the first sentence “be able to” into Turkish either as *yapabiliyor* (present continuous tense) or *yapabilir* (present tense). In NMT, the tense consistency observed between the verbs is maintained in the second verb translation *sunabiliyor* (present continuous tense) but not in the translation of the last verb *yorumlamak* (infinitive). Similarly, an inconsistency is noted in SMT-2 and SMT-3 outputs.

In addition, the translation of sentence “collect and present data in the experiment between different metals and metal's solutions in numerical and visual forms” that is produced by the SMT-2 includes an ambiguity on whether “the experiments are in various forms numerically or visually” or “the data to be collected and presented is in numerical and visual forms”. In this sense, it can be claimed that the output of SMT-2 may lead to

some inaccuracy and ambiguity. Therefore, it may not be wrong to claim that this sentence definitely requires post-editing and it should be evaluated within the scope of the fluency issue of the MQM.

Example 16:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. select and apply mathematical problem-solving techniques to recognize correct patterns in explaining change the number of bacteria for each day or hour ii. describe patterns as relationships or general rules consistent with correct equation that you get at the end of the pattern iii. verify whether patterns work for the second bacteria C example (Mathematics – Criterion B- Year 5)	Öğrenci şunları yapabiliyor: i. Her gün veya saat için bakteri sayısını değiştirmeyi açıklarken doğru desenleri tanımak için matematiksel problem çözme tekniklerini seçmek ve uygulamak ii. desenleri, desenin sonunda elde ettiğiniz doğru denklemlerle tutarlı ilişkiler veya genel kurallar olarak tanımlamak iii. desenlerin ikinci bakteri C örneği için çalışıp çalışmadığını doğrulamak	Öğrenci şunları yapabilir: ben. Her gün veya saat için bakteri sayısını değiştirmeyi açıklamada doğru kalıpları tanımak için matematiksel problem çözme tekniklerini seçin ve uygulayın ii. Kalıpları, kalıbın sonunda elde ettiğiniz doğru denklemlerle tutarlı ilişkiler veya genel kurallar olarak tanımlayın iii. kalıpların ikinci bakteri C örneği için çalışıp çalışmadığını doğrulayın	Öğrenci şunları yapabilir: i. her gün veya saat için bakteri sayısındaki değişimi açıklarken doğru kalıpları tanımak için matematiksel problem çözme tekniklerini seçer ve uygular ii. kalıpları, desenin sonunda elde ettiğiniz doğru denklemlerle tutarlı ilişkiler veya genel kurallar olarak tanımlayın iii. kalıpların ikinci bakteri C örneği için çalışıp çalışmadığını doğrulayın

Comment:

In this example, the expression “change” is translated in two different ways: *değiştirme* and *değişim*. The former equivalent is used in NMT and SMT-1 while the latter is applied in the SMT-2. Within this context, it would be more appropriate to use *değiştirme*. In addition, the word “pattern” is rendered as *desen* and *kalıp* in three of the MT. In

accordance with the context, these may not be proper translations. Instead, Turkish equivalents such as *örüntü*, *motif* could be more suitable for the context.

Moreover, number format issues are seen both in SMT-1 and SMT-2 in this example. In SMT-1, the number “i.” is translated as a pronoun while numbers “ii.” and “iii.” are translated as letters in SMT-2. Both are problematic outputs in terms of locale-convention issue. Even though all target texts are fluent, the imperative grammar structure that is used in SMT-1 and SMT-2 might not be the most convenient grammatical form to apply.

Example 17:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. state a problem to be tested by a scientific investigation in order to build a bridge in Istanbul. ii. outline a hypothesis to design a bridge considering the properties. iii. based on your hypothesis outline the variables iv. design a method to reach the aim with limited resources (<i>Sciences Criterion B – Year 5</i>)	Öğrenci şunları yapabiliyor: i. İstanbul'da bir köprü inşa etmek için bilimsel bir soruşturma ile test edilecek bir sorunu belirtmek ii. özellikleri göz önünde bulundurarak bir köprü tasarlamak için bir hipotez belirlemek iii. hipotezinize dayanarak değişkenleri özetlemek iv. sınırlı kaynaklarla amaca ulaşmak için bir yöntem tasarlamak	Öğrenci şunları yapabilir: ben. İstanbul'da bir köprü inşa etmek için bilimsel bir araştırmayla test edilecek bir problem belirtin. ii. özellikleri göz önünde bulundurarak bir köprü tasarlamak için bir hipotezin ana hatlarını çizer. iii. hipotezinize dayanarak değişkenleri ana hatlarıyla belirtin iv. sınırlı kaynaklarla amaca ulaşmak için bir yöntem tasarlamak	Öğrenci şunları yapabilir: i. İstanbul'da bir köprü inşa etmek için bilimsel bir araştırma ile sınanacak bir sorunu belirtiniz. ii. özellikleri göz önünde bulundurarak bir köprü tasarlamak için bir hipotezin ana hatlarını çizer. iii. hipotezinize dayanarak değişkenleri özetleyin iv. sınırlı kaynaklarla hedefe ulaşmak için bir yöntem tasarlamak

Comment:

In this example, there are issues regarding the number format in SMT-1 and SMT-2. In SMT-1, number “i.” is mistranslated as a pronoun while in SMT-2, numbers “ii.” and “iii.” are translated as letters instead of numbers. These problems are dealt with the locale convention issue of the MQM.

It can be claimed that three of the machine translation outputs are intelligible and accurate. The translation produced by NMT has a consistent grammatical structure whereas the syntactic structure of SMT-1 and SMT-2 is same and as follows: imperative, present tense, imperative, infinitive. It has a negative impact on the fluency of the target text.

Example 18:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. identify the relevant elements of city planning ii. select adequate mathematical strategies to identify whether a two-way road has car parking area in both sides of the road or not iii. apply the selected mathematical strategies to identify whether a two-way road has car parking area in both sides of the road not, by reaching a valid solution iv. describe the degree of accuracy of the solution of identifying whether a two-way road has car parking area in both sides of the road not v. state correctly whether the solution makes sense in the context	Öğrenci şunları yapabiliyor: i. şehir planlamasının ilgili unsurlarını belirlemek ii. iki yönlü bir yolun her iki tarafında otopark alanı olup olmadığını belirlemek için yeterli matematiksel stratejileri seçmek iii. geçerli bir çözüme ulaşarak, iki yönlü bir yolun her iki tarafında da otopark alanı olup olmadığını belirlemek için seçilen matematiksel stratejileri uygulamak iv. İki yönlü bir yolun her iki tarafında da otopark alanı olup olmadığını belirleme çözümünün doğruluk derecesini tanımlamak	Öğrenci şunları yapabilir: ben. şehir planlamasının ilgili unsurlarını belirlemek ii. iki yönlü bir yolun her iki tarafında da otopark alanı olup olmadığını belirlemek için yeterli matematiksel stratejileri seçin. iii. Geçerli bir çözüme ulaşarak, iki yönlü bir yolun her iki tarafında da otopark alanı olup olmadığını belirlemek için seçilen matematiksel stratejileri uygulayın. iv. İki yönlü bir yolun, yolun her iki tarafında da otopark alanı olup olmadığını belirleme çözümünün doğruluk derecesini tanımlayın.	Öğrenci şunları yapabilir: i. şehir planlamasının ilgili unsurlarını belirlemek ii. iki yönlü bir yolun yolun her iki tarafında otopark alanı olup olmadığını belirlemek için yeterli matematiksel stratejileri seçin. iii. geçerli bir çözüme ulaşarak, iki yönlü bir yolun yolun her iki tarafında otopark alanı olup olmadığını belirlemek için seçilen matematiksel stratejileri uygulayın iv. iki yönlü bir yolun yolun her iki tarafında otopark alanı olup olmadığını belirleme çözümünün doğruluk derecesini açıklayınız.

of the authentic real-life situation (<i>Mathematics – Criterion D – Year 5</i>)	v. çözümün gerçek yaşam durumu bağlamında anlamlı olup olmadığını doğru bir şekilde belirtmek	v. çözümün otantik gerçek yaşam durumu bağlamında anlamlı olup olmadığını doğru bir şekilde belirtin	v. çözümün gerçek hayattaki gerçek durum bağlamında mantıklı olup olmadığını doğru bir şekilde belirtin
---	---	--	---

Comment:

In the given example, there is a mistake related to the number format in the SMT-1 output. SMT-1 conveyed the number “i.” as *ben* which is the equivalent of first-person singular pronoun in Turkish. It is a repeated mistake and it is observed in all SMT-1 outputs. Hence, it can be concluded that SMT-1 has a problematic issue in regards to the translation of numbers.

In terms of grammatical structure used in NMT, it can be expressed that it is consistent and fluent. As to SMT-1, sentence numbered i. is in infinitive form while the rest of them are in imperative structure. Whereas the syntactic structure of SMT-2 is as follows: infinitive, imperative, imperative (formal voice), imperative.

Last but not least, the expression “authentic real-life situation” is translated in NMT as *gerçek yaşam durumu*; in SMT-1 as *otantik gerçek yaşam durumu*; in SMT-2 as *gerçek hayattaki gerçek durum*. Of these three statements, the first one sounds more natural in Turkish.

Example 19:

Source Text	NMT	SMT-1	SMT-2
Students: i. identify some relevant elements of triangle figure and similarity relation. ii. select some appropriate mathematical strategies when solving market	Öğrenciler: i. üçgen şekil ve benzerlik ilişkisinin bazı ilgili unsurlarını tanımlar. ii. pazar mesafesi problemini çözerken bazı uygun	öğrenciler: ben. üçgen şekli ve benzerlik ilişkisinin bazı ilgili unsurlarını tanımlar. ii. Pazar mesafesi problemini çözerken bazı uygun	Öğrenciler: i. Üçgen şekil ve benzerlik ilişkisinin ilgili bazı unsurlarını tanımlar. ii. piyasa mesafesi problemini çözerken bazı uygun

distance problem. iii. apply some selected mathematical strategies successfully to get similarity equation. iv. justify some degree of accuracy of a solution. v. justify whether the using and equation make sense in the context of the market problem. (Mathematics – Criterion D – Year 1)	matematiksel stratejiler seçer. iii. benzerlik denklemini elde etmek için seçilen bazı matematiksel stratejileri başarıyla uygular. iv. bir çözümün bir dereceye kadar doğruluğunu haklı çıkarır. v. kullanma ve denklemin piyasa sorunu bağlamında anlamlı olup olmadığını haklı çıkarır.	matematiksel stratejiler seçin. iii. benzerlik denklemini elde etmek için seçilen bazı matematiksel stratejileri başarıyla uygular. iv. bir çözümün bir dereceye kadar doğruluğunu gerekçelendirin. v. Kullanım ve denklemin piyasa sorunu bağlamında anlamlı olup olmadığını gerekçelendirin.	matematiksel stratejileri seçin. iii. benzerlik denklemini elde etmek için seçilen bazı matematiksel stratejileri başarıyla uygular. iv. bir çözümün bir dereceye kadar doğruluğunu haklı göstermek. v. piyasa problemi bağlamında kullanımın ve denklemin anlamlı olup olmadığını gerekçelendirmek.
--	---	---	---

Comment:

In the SMT-1 of this example, a capitalization mistake is detected together with a number format issue in which the number “i.” is translated as the pronoun *ben*. Likewise, two numbers in SMT-2 are translated as letters. These mistranslations are evaluated within the scope of locale-convention of the MQM.

Another issue is the grammatical structure of target texts. In NMT, all sentences are accurately rendered in present tense. However, SMT outputs show random grammatical structure in the sentences. In SMT-1, sentences are given in present tense, imperative, present tense, imperative and imperative whereas they are translated in the following forms in SMT-2: present tense, imperative, present tense, infinitive, infinitive. These are hurdles against the fluency and readability of the target texts.

The word “justify” in the source sentence numbered v. is translated as *haklı çıkarmak* in NMT. However, this expression is translated as *gerekçelendirmek* generally in this context. In SMT-1, it is translated accordingly. As to the SMT-2, both of the translations are used in the target text.

Example 20:

Source Text	NMT	SMT-1	SMT-2
The student is able to: i. use some appropriate mathematical and academic language while explaining mathematical and logical steps to identify innocent and guilty people in an investigation of robbery ii. use appropriate forms and symbols of mathematical representation to identify innocent and guilty people in an investigation of robbery adequately iii. communicate through lines of reasoning that are able to be understood for identifying innocent and guilty people in an investigation of robbery, although these are not always coherent iv. adequately organize information using a logical structure to identify innocent and guilty people in an investigation of robbery (Mathematics –	Öğrenci şunları yapabiliyor: i. bir soygun soruşturmasında masum ve suçluları tespit etmek için matematiksel ve mantıksal adımları açıklarken bazı uygun matematiksel ve akademik dil kullanmak ii. soygun soruşturmasında masum ve suçlu kişileri tespit etmek için matematiksel temsilin uygun formlarını ve sembollerini kullanmak iii. her zaman tutarlı olmasa da, bir soygun soruşturmasında masum ve suçlu insanları tespit etmek için anlaşılabilen akıl yürütme hatları aracılığıyla iletişim kurmak iv. bir soygun soruşturmasında masum ve suçlu kişileri tespit etmek için mantıklı bir yapı kullanarak bilgileri yeterince organize etmek	Öğrenci şunları yapabilir: ben. Bir soygun soruşturmasında masum ve suçlu kişileri belirlemek için matematiksel ve mantıksal adımları açıklarken uygun matematiksel ve akademik bir dil kullanın ii. Bir soygun soruşturmasında masum ve suçlu kişileri yeterince tanımlamak için matematiksel temsilin uygun biçimlerini ve sembollerini kullanın iii. bir soygun soruşturmasında masum ve suçlu kişileri tespit etmek için anlaşılabilir muhakeme çizgileri aracılığıyla iletişim kurar, ancak bunlar her zaman tutarlı değildir. iv. bir soygun soruşturmasında masum ve suçlu kişileri belirlemek için mantıksal bir yapı kullanarak bilgileri yeterince organize edin	Öğrenci şunları yapabilir: i. soygun soruşturmasında masum ve suçlu kişileri tanımlamak için matematiksel ve mantıksal adımları açıklarken uygun matematiksel ve akademik bir dil kullanın ii. soygun soruşturmasında masum ve suçlu kişileri yeterince tanımlamak için uygun matematiksel temsil biçimlerini ve sembollerini kullanın iii. bir soygun soruşturmasında masum ve suçlu kişileri tespit etmek için anlaşılabilir akıl yürütme çizgileri aracılığıyla iletişim kurun, ancak bunlar her zaman tutarlı değildir iv. soygun soruşturmasında masum ve suçlu kişileri tespit etmek için mantıksal bir yapı kullanarak bilgiyi yeterince organize etmek

Criterion C – Year 5)			
--------------------------	--	--	--

Comment:

In this example, all of the target texts are accurately translated. Only in SMT-1, the number “i.” is translated as a pronoun. In addition, the expression “lines of reasoning” is respectively translated as: *akıl yürütme hatları*, *muhakeme çizgileri*, *akıl yürütme çizgileri*. As it can be seen through the examples, these are literal translations. Instead, it could have been translated as *akıl yürütme* or *muhakeme* in order to enhance intelligibility of the sentence. Only in NMT, a consistent use of syntactic structure is observed. Therefore, it is more fluent than the other target texts. Apart from these points, machine translation outputs are generally accurate.

8.1 NMT and SMT Output Analysis Evaluation

Various NMT and SMT translations are examined in line with the Multidimensional Quality Metrics in the previous section. Upon these examinations, it can be concluded that a domain-specific customized neural machine translation system may have a tendency to produce more fluent, accurate and terminologically consistent target texts. It should be noted that generic SMT systems also put forth quality outputs in terms of overall accuracy. However, some problems especially in consistency within the scope of fluency and in locale convention occurred continuously. In addition, mistranslations or untranslated parts are found in some SMT produced target texts. Taking the above given detailed comments and analyses into consideration, it can be claimed that a neural machine translation engine customized domain-specific corpus is able to outperform generic statistical machine translation systems in regards to fluency and ability to produce texts suitable for the audience and context i.e. locale convention.

8.2 BLUE Score of the Compared NMT and SMT Outputs

In this part, BLEU Scores of the previously given NMT and SMT comparisons are calculated to check the score differences. Since different machine translation outputs are used in the comparisons, a built-in BLEU Score could not be used in this process. Therefore, a BLUE Score metric system required to be created via Python. To this end, a professional help is received from a code-writer. With this end, open-source codes for

BLUE Score metric system are researched. At first, the codes obtained via forums on the Internet did not work. After a couple of trying, one of the code worked. The code that worked is shared below:

```
nltk download
command line: pip install nltk
import nltk
python come
nltk.download()
from nltk.translate.bleu_score import sentence_bleu
reference = [['this', 'is', 'a', 'test'], ['this', 'is', 'test']]
candidate = ['this', 'is', 'a', 'test']
score = sentence_bleu(reference, candidate)
print(score)
```

After codes are written, each output i.e. candidate translation is compared with the reference text in the program. To compare the sentences, first the following command is written:

```
reference = [['this', 'is', 'a', 'test'], ['this', 'is', 'test']]
```

Each word is written between apostrophes and a comma is added after each word in apostrophe. A space must be left after the comma. If any apostrophe or comma lack, the system fails. And it requires to be repeated from the very beginning. Therefore, this process must be done delicately. At the beginning and end of the reference sentence, two square brackets must be put. After reference sentence is finished, candidate translation is written as follows:

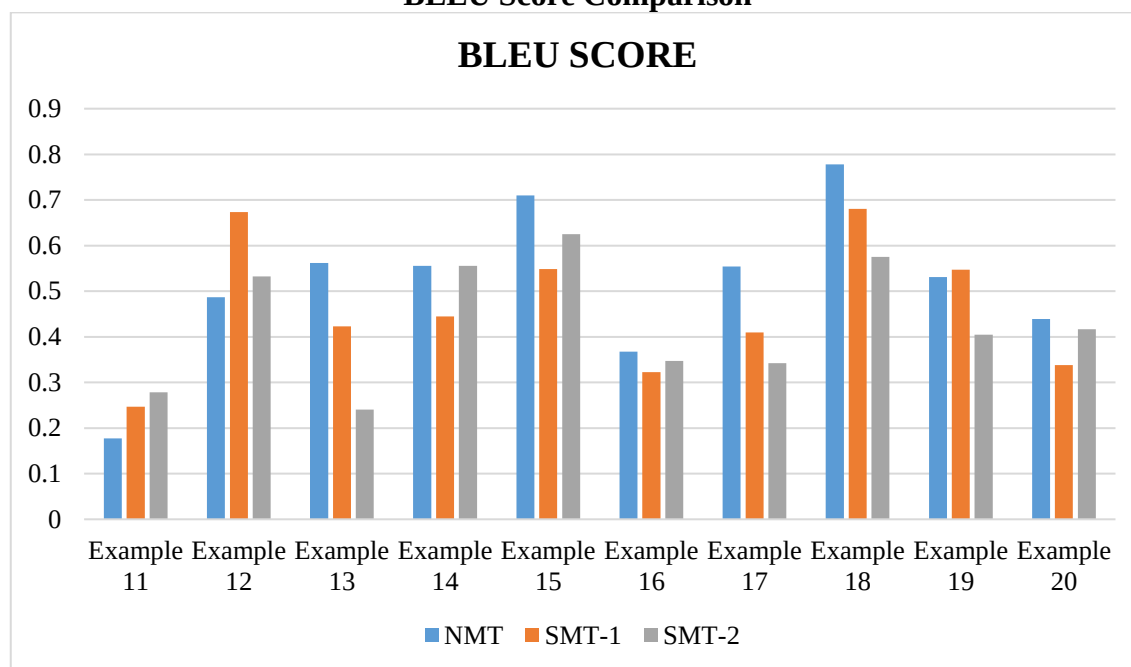
```
candidate = ['this', 'is', 'a', 'test']
```

Like in the reference command, each word is written between apostrophes and commas in candidate sentence, as well. The only difference is that one square bracket is added at the beginning and end of the sentence. After this process is completed, the following commands are written respectively:

```
score = sentence_bleu(reference, candidate)
print(score)
```

Finally, the results are obtained and written in the tables (see Appendix II). Below is a graphic showing the all of the results acquired at the end of this process.

Table 2
BLEU Score Comparison



In accordance with the data shown on the graphic, it can be stated that the performance of NMT generally surpasses the other two SMT performances. In 6 of the examples, BLEU Score of the NMT is the highest. SMT-1 has the highest score in 2 examples while SMT-2 has the highest score only in 1 example and it has an equal score with NMT in another example.

Before the BLEU Score evaluation, examples were analyzed through Multidimensional Quality Metrics. It has shown that the NMT output quality is higher than the SMT output quality. In this section, examples are calculated via BLUE Score evaluation metric system. In light of this metric evaluation, it can be concluded that the customized neural machine translation engine i.e. Microsoft Custom Translator has proved itself as a better option to generic SMT systems in regard to the area-specific translation.

8.3 Examination on NMT and SMT Outputs for Further Training

A comparative analysis and the BLUE Score metric results are described elaborately in the above given sections. In this part, a table listing the mistakes encountered in all three of the MT outputs is provided below.

Table 3
NMT and SMT Output Error List

	NMT	SMT-1	SMT-2
Example 11		number format	number format
	inconsistent terminology with domain		
	inconsistent syntactic structure	inconsistent syntactic structure	inconsistent syntactic structure
Example 12		number format	
	inconsistent terminology with domain	inconsistent terminology with domain	inconsistent terminology with domain
Example 13		number format	number format
	inaccurate translation		
			untranslated
Example 14		number format	
Example 15		number format	
	inconsistent syntactic structure	inconsistent syntactic structure	inconsistent syntactic structure
			ambiguity
Example 16		number format	number format
		inconsistent syntactic structure	inconsistent syntactic structure
	inconsistent terminology with domain	inconsistent terminology with domain	inconsistent terminology with domain
Example 17		number format	number format
		inconsistent syntactic structure	inconsistent syntactic structure
Example 18		number format	
		inconsistent syntactic structure	inconsistent syntactic structure
		literal translation	literal translation
Example 19		capitalization	
		number format	number format
		inconsistent syntactic structure	inconsistent syntactic structure
Example 20		number format	

		inconsistent syntactic structure	inconsistent syntactic structure
--	--	----------------------------------	----------------------------------

On the given table, all of the errors and mistakes are listed. As it can be seen, all of them has some problems in terms of terminology use. SMT-1 has major issues with number format together with SMT-2. Inconsistent syntactic structure is a problem observed mainly in SMT outputs. Besides, there are less recurring mistakes such as literal translation and ambiguity.

In light of this information, the NMT engine could be trained with further data to increase the translation quality. With this aim, summative assessment criteria that are applied in actual classroom environments can also be uploaded to the custom engine in order to train the system with relevant course terminologies. In so doing, the custom NMT may put forth outputs with higher accuracy and consistency rates.

CONCLUSION

The main objective of this thesis was to provide accurate, consistent and fluent translations of the summative assessment criteria that are produced within the scope of the International Baccalaureate Middle Years Programme in a shorter period of time. These criteria have a significant role in the grade assessment of the students receiving MYP education. However, there is a continuous and intense volume of translation need of these documents in some schools offering MYP in Turkey. Hence, it was claimed that a custom NMT engine could be able to offer a solution to the urgent and continuous translation need of the MYP summative assessment criteria.

To that end, the concepts of machine translation, neural machine translation and customization were explained meticulously. Then, the IB was introduced together with the MYP and other IB education programmes. Within this scope, summative assessment criteria and their role in the education programme was described. In this regard, IB MYP English-Turkish summative assessment criteria were collected one by one and editing steps were taken when necessary. Afterwards, these files were aligned by using LF Aligner, a stand-alone alignment tool. Each stage of this process was documented in the thesis.

As the next step, Microsoft Custom Translator was decided to be used among various other NMT tools and the license procedures were completed. In the Custom Translator, a project was created and previously aligned training data and other relevant documents were uploaded into that project. Later, a model was created within the project and the training process of the model was started. At the end of the customization process, the model was deployed by selecting one or more region. In this case, only Europe region was selected. Each of these steps were also included and described in the thesis together with screenshots. At the end of these processes, the deployed model needed to be used in production.

In order to do this, the system needed to be run by some basic codes. Since this step required more expert knowledge in the field of software and computational skills, a professional help was received from an expert. Necessary programmes were installed and required codes were written into these programmes with the help of the software expert. Finally, the model was ready to be used for translation. With this aim, test materials were

obtained from an IB Continuum School based in Istanbul, Turkey. These test materials were used in actual MYP offering school environments. Several examples were received and uploaded to the system and translated. Later, these NMT outputs were analyzed in Chapters 7 and 8 of this thesis. As to this analysis process, two analysis and evaluation system were applied: Multidimensional Quality Metrics and BLEU Score evaluation metric system. In the analysis of the examples, each was provided with detailed comments that were written with respect to MQM Core issues. Then, the baseline BLEU Score and the BLEU score of the model that was produced within the Custom Translator was provided.

In order to have a better grasp of NMT output quality, the NMT outputs were compared with the outputs of some other generic SMT systems. To this end, different examples were acquired as test data from the previously mentioned IB Continuum School based in Istanbul, Turkey. These examples were analyzed in accordance with the MQM metrics as well as a detail BLEU Score evaluation. In this stage, an open-source code was obtained from the Internet and was written with the help of a software specialist to measure and evaluate the BLEU Score of the NMT and SMT outputs. The results of this evaluation were provided in a graphic in Chapter 8. At the end of these processes, it was proved that a customized NMT engine is able to produce more accurate, consistent, fluent, high-quality output than generic SMT systems.

In consequence, it can be claimed that this thesis humbly put forth that a neural machine translation system which is trained with quality data is able to generate proper, accurate and consistent translations in a short period of time. However, it should be noted that more can be done for the custom NMT system to produce more quality target texts. For instance, criteria documents that are used in real classroom environments can be obtained and uploaded to the engine to increase the output quality. In addition, the scope of the research can be expanded to other MYP document translation or more general IB documents. In such a case, since the train materials will increase in number, the target text quality can also improve.

It is a fact that the study was conducted with limited resources and only in English-Turkish language pair due to time and scope constraints. However, it is hoped that this thesis may help translators, educators and anyone working in this field. It is also aspired

that this thesis may inspire other researchers to make further studies on the NMT technology.



REFERENCES

- "About the IB." *International Baccalaureate*. n.d. <https://www.ibo.org/about-the-ib/> (accessed April 4, 2021).
- ALS Staff. "Why Do You Need Technology in Translation?" *Affordable Language Services*. July 9, 2019. <https://www.affordablelanguageservices.com/why-do-you-need-technology-in-translation/> (accessed June 1, 2021).
- Brockmann, Daniel. "SDL Machine Translation: The Future of Neural Machine Translation Is Here." *Trados*. November 28, 2019. <https://www.sdltrados.com/blog/future-nmt.html> (accessed April 4, 2021).
- Castilho, Sheila , Joss Moorkens, Federico Gaspari, Rico Sennrich, Andy Way, and Panayota Georgakopoulou. "Evaluating MT for Massive Open Online Courses: A Multifaceted Comparison between PBSMT and NMT Systems." *Machine Translation*, no. 32 (2018): 255-278.
- Cho, Kyunghyun , Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches." *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, 25 October 2014. 103-111.
- Dell, Christopher. "Why Use A Termbase or Glossary When Translating An App?" *Locale*. n.d. <https://www.locale.to/why-use-a-termbase-or-glossary-when-translating-an-app/> (accessed August 30, 2021).
- Doğru, Gökhan. "Statistical Machine Translation Customization between Turkish and 11 Languages." *transLogos Translation Studies Journal*, 3:1 (2020): 98-121.
- Doherty, Stephen. "The Impact of Translation Technologies." *International Journal of Communication*, 2016: 947–969.
- "Evaluating Models." *Google Cloud*. n.d. <https://cloud.google.com/translate/automl/docs/evaluate> (accessed June 3, 2021).
- Forcada, Mikel L. "Making Sense of Neural Machine Translation." *Translation Spaces* 6, no. 2 (2017): 291-309.
- Ghislandi, Massimo. "What Is CAT Tool?" *SDL Trados*. n.d. <https://www.sdltrados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html> (accessed April 4, 2021).
- Holmes , James S. "The Name and Nature of Translation Studies." In *Translated!: Papers on Literary Translation and Translation Studies*, by James S. Holmes, 66-80. Amsterdam: Rodopi, 1988.

- Hutchins, W. John. "Machine Translation: A Brief History." In *Concise History of the Language Sciences From the Sumerians to the Cognitivists*, by E. F. K. Koerner and R. E. Asher, 431-445. Oxford: Pergamon Press, 1995.
- "Is Node JS a Framework? Learn About the Popular Platform." *Effectus Software*. n.d. <https://effectussoftware.com/blog/node-js-a-framework/> (accessed May 23, 2021).
- Lavie, Alon. "Evaluating the Output of Machine Translation Systems." *AMTA*. Xiamen, China, September 19, 2011.
- Lommel, Arle Richard , Aljoscha Burchardt, and Hans Uszkoreit. *Multidimensional Quality Metrics : A Flexible System for Assessing Translation Quality*. 2013.
- Lommel, Arle, Aljoscha Burchardt, and Hans Uszkoreit. *Multidimensional Quality Metrics (MQM) Definition (1.0)*. 12 30, 2015. <https://www.qt21.eu/mqm-definition/definition-2015-12-30.html#scoring-algorithm> (accessed August 29, 2021).
- Mandal, Some Aditya. "Evolution of Machine Translation." *Towards Data Science*. June 4, 2019. <https://towardsdatascience.com/evolution-of-machine-translation-5524f1c88b25> (accessed June 3, 2021).
- "MYP Assessment." *Istanbul International Community School*. n.d. <https://www.iics.k12.tr/academics/myp-assessment/> (accessed May 23, 2021).
- "Neural Machine Translation: Embarking on the New Frontier." *Ulatus*. April 16, 2018. <https://www.ulatus.com/translation-blog/neural-machine-translation-embarking-on-the-new-frontier/> (accessed September 1, 2021).
- Node.js*. n.d. <https://nodejs.org/en/> (accessed May 23, 2021).
- Öner Bulut, Senem. "Integrating Machine Translation into Translator Training: Towards 'Human Translator Competence'?" *transLogos Translation Studies Journal*, 2:2 (2019): 1-26.
- Öner, Işın. "İnsan Çevirisi ve Makine Çevirisi: Rekabet mi İşbirliği mi?" Istanbul: Arel University, 26 November 2018.
- Öner, Işın, and Senem Öner Bulut. "Post-Editing Oriented Human Quality Evaluation of Neural Machine Translation in Translator Training: A Study on Perceived Difficulties and Benefits." *transLogos Translation Studies Journal*, 4:1 (2021): 100-124.
- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "Bleu: A Method for Automatic Evaluation of Machine Translation." *ACL 2002: Proceedings of the*

40th Annual Meeting on Association for Computational Linguistics.
Philadelphia, USA: Association for Computational Linguistics, 2002. 311-318.

"Programmes." *International Baccalaureate*. n.d. <https://www.ibo.org/programmes/> (accessed April 12, 2021).

"Teaching the IB in Your Language." *International Baccalaureate*. n.d. <https://www.ibo.org/about-the-ib/the-ib-by-country/t/turkey/> (accessed April 16, 2021).

"The IB by Country." *International Baccalaureate*. n.d. <https://www.ibo.org/about-the-ib/the-ib-by-country/t/turkey/> (accessed April 12, 2021).

Toury, Gideon. *Descriptive Translation Studies and Beyond*. John Benjamins Publishing, 1995.

"Train A Model." *Microsoft*. August 17, 2020. <https://docs.microsoft.com/en-us/azure/cognitive-services/translator/custom-translator/how-to-train-model> (accessed February 28, 2021).

"Translation 101: What Is A CAT Tool?" *Trados*. n.d. <https://www.trados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html> (accessed April 4, 2021).

"Why Did We Build Visual Studio Code?" *Visual Studio Code*. May 5, 2021. <https://code.visualstudio.com/docs/editor/whyvscode> (accessed May 23, 2021).

Training Material References:

International Baccalaureate Organization[®]. May, 2014. *Arts Guide, 2014*. United Kingdom. [Online] Available at: https://www.fultonschools.org/cms/lib/GA50000114/Centricity/Domain/220/MYP_Arts_Guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Design Guide, 2014*. United Kingdom. [Online] Available at: <http://www.kis.in/wp-content/uploads/2015/07/Design-Guide.pdf> (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Individuals and Societies Guide, 2014*. United Kingdom. [Online] Available at: https://www.spps.org/site/handlers/filedownload.ashx?moduleinstanceid=38367&dataid=21204&FileName=individuals_and_societies_guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Language Acquisition Guide, 2014*. United Kingdom. [Online] Available at:

https://www.spps.org/site/handlers/filedownload.ashx?moduleinstanceid=38371&dataid=21214&FileName=language_acquisition_guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Language and Literature Guide, 2014*. United Kingdom. [Online] Available at: https://www.spps.org/site/handlers/filedownload.ashx?moduleinstanceid=38369&dataid=21208&FileName=language_and_literature_guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Mathematics Guide, 2014*. United Kingdom. [Online] Available at: https://lcismyp.files.wordpress.com/2018/05/math_guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Physical and Health Education Guide, 2014*. United Kingdom. [Online] Available at: https://www.spps.org/site/handlers/filedownload.ashx?moduleinstanceid=38375&dataid=21229&FileName=p_e_health_guide_2014.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Projects Guide, 2014*. United Kingdom. [Online] Available at: https://os-mgubec.eu/userfiles/Dokumenti/IB_Projects_Guide.pdf (accessed April 4, 2021)

International Baccalaureate Organization[®]. May 2014. *Science Guide, 2014*. United Kingdom. [Online] Available at: https://www.spps.org/site/handlers/filedownload.ashx?moduleinstanceid=38378&dataid=21234&FileName=sciences_guide_2014.pdf (accessed April 4, 2021)

APPENDICES

APPENDIX I: BLEU SCORE OF MT OUTPUT EXAMPLES

Example 11	
	BLEU Score
NMT	0.17714475283946735
SMT-1	0.24684653565911077
SMT-2	0.27815405049907466

Example 12	
	BLEU Score
NMT	0.48656669832987026
SMT-1	0.6730933670690624
SMT-2	0.5322771566654712

Example 13	
	BLEU Score
NMT	0.5619716962400328
SMT-1	0.42291922939267523
SMT-2	0.24067683805330567

Example 14	
	BLEU Score
NMT	0.5555555555555556
SMT-1	0.4444444444444444
SMT-2	0.5555555555555556

Example 15	
	BLEU Score

NMT	0.7096774193548387
SMT-1	0.5483870967741935
SMT-2	0.625

Example 16	
	BLEU Score
NMT	0.3671510216651532
SMT-1	0.3226080730683482
SMT-2	0.3474411871253603

Example 17	
	BLEU Score
NMT	0.5542162712489429
SMT-1	0.4092079294124895
SMT-2	0.34222297297205356

Example 18	
	BLEU Score
NMT	0.7781962810712585
SMT-1	0.6801379194102606
SMT-2	0.5754525504582915

Example 19	
	BLEU Score
NMT	0.5312369676130652
SMT-1	0.547249873088654
SMT-2	0.40479778424280616

Example 20	
-------------------	--

	BLEU Score
NMT	0.4389167617930115
SMT-1	0.3378755653407851
SMT-2	0.41656763991131696



ÖZGEÇMİŞ			
Adı, Soyadı	Merve		ARAZ
Doğum Yeri ve Yılı			
Bildiği Yabancı Diller	İngilizce		Arapça
Yetkinlik Düzeyi	İleri		Orta
Eğitim Durumu	Başlama – Bitirme Yılı		Kurum Adı
Lise	2010	2014	Dr. Kemal Naci Ekşi Anadolu Lisesi
Lisans	2014	2018	İstanbul 29 Mayıs Üniversitesi
Yüksek Lisans	2018	2021	İstanbul 29 Mayıs Üniversitesi
Doktora			
Çalıştığı Kurum(lar)	Başlama - Ayrılma Yılı		Çalışılan Kurumun Adı
3.	2020	Devam ediyor	NUN Okulları
Üye Olduğu Bilimsel ve Mesleki Kuruluşlar			
Katıldığı Proje ve Toplantılar		1. Avrupa Birliği Başkanlığı 2018 Genç Çevirmenler Yarışması	
Yayınlar:			
Diğer:			
İletişim (e-posta):			
Tarih		14/09/2021	
İmza			
Adı Soyadı		Merve ARAZ	

