UNIVERSITY OF ECONOMICS IN BRATISLAVA

FACULTY OF APPLIED LANGUAGES

Registration number: 106006/M/2019/36069194042589444

USE OF MACHINE TRANSLATION TOOLS IN COMMERCIAL TRANSLATION IN SLOVAKIA

Diploma thesis

2019 **Bc. Jakub Hanúsek**

UNIVERSITY OF ECONOMICS IN BRATISLAVA

FACULTY OF APPLIED LANGUAGES

Use of Machine Translation Tools in Commercial Translation in Slovakia

Diploma thesis

Study programme: Foreign Languages and Intercultural Communication

Field of study: Foreign Languages and Cultures

Consultation centre: Department of Linguistics and Translatology

Tutor: Mgr. Jozef Štefčík, PhD.

Bratislava 2019 **Bc. Jakub Hanúsek**

Affirmation

I hereby affirm that this thesis represents my own original research and writing and that I have referenced all appropriate source materials.

Date: 04.04.2019 ……...

Acknowledgments

I would like to thank Mgr. Jozef Štefčík, PhD. for his assistance in the researching and writing of this paper, as well as for his guidance and support.

ABSTRAKT

HANÚSEK, Jakub: *Využitie nástrojov strojového prekladu v oblasti komerčného prekladu na Slovensku*. – Ekonomická Univerzita v Bratislave. Fakulta aplikovaných jazykov; Katedra jazykovedy a translatológie. – Vedúci diplomovej práce: Mgr. Jozef Štefčík, PhD. – Bratislava: FAJ EU, 2019, 90 s.

Stanoveným cieľom tejto diplomovej práce je miera využitia strojového prekladu v komerčnom preklade na Slovensku. Zameriame sa na históriu strojového prekladu až po súčasnosť. Pokúsime sa určiť a kvantifikovať využitie strojového prekladu prekladateľskými subjektami na Slovensku.

Diplomová práca je rozdelená do troch častí. V prvej časti, ktorá je venovaná teoretickým znalostiam o danej problematike si zhrnieme históriu strojového prekladu. Následne sa budeme sústrediť na charakteristiku a metódy rôznych prístupov strojového prekladu. Pokúsime sa recipientovi vysvetliť pojmy ako neurónový preklad, pre-editáciu a post-editáciu strojového prekladu a mnohé ďalšie.

Druhá časť našej diplomovej práce obsahuje praktickú časť. V praktickej časti budeme analyzovať dotazník, ktorý sme distribuovali medzi prekladateľov. Pomocou dotazníka sa budeme snažiť zodpovedať vytvorené hypotézy a určiť tak stav prekladateľského priemyslu na Slovensku.

V tretej časti našej diplomovej práce zhrnieme výsledky dotazníka a skonštatujeme, či sa nám podarilo splniť stanovený cieľ našej diplomovej práce. V tejto časti tak isto zdôrazníme rôzne možnosti vývoja prekladateľského priemyslu na Slovensku a budeme sa sústrediť aj na možné odporúčania do budúcnosti.

Kľúčové slová: história, preklad, strojový preklad, neurónový preklad, história strojového prekladu, budúcnosť strojového prekladu, Google Prekladač, SYSTRAN

ABSTRACT

HANÚSEK, Jakub: *Use of machine translation tools in commercial translation in Slovakia*. – University of Economics in Bratislava. Faculty of Applied Languages; Department of Linguistics and Translation. – Tutor: Mgr. Jozef Štefčík, PhD. – Bratislava: FAJ EU, 2019, 90 p.

The fixed aim of this thesis is the extent of use of the Machine Translation in commercial translation in Slovakia. We will focus on the history of the Machine Translation up to the present. We will try to determine and quantify the use of the Machine Translation by translation entities in Slovakia.

The diploma thesis is divided into three parts. In the first part, which contains the theoretical knowledge of this field, we summarise the history of the Machine Translation. Later, we will focus on the characteristics and methods of the different Machine Translation approaches. We will try to explain to the recipient terms such as neural translation, pre-editing and post-editing of the Machine Translation and many others.

The second part of our diploma thesis includes a practical part. In the practical part, we will analyse a questionnaire, which we distributed among translators. Using the questionnaire, we will try to answer several created hypotheses and determine the state of the translation industry in Slovakia.

In the third part of our diploma thesis, we summarise the results of the questionnaire and state whether we managed to meet the fixed aim of our diploma thesis. In this section, we will also highlight the various developments in the translation industry in Slovakia and focus on possible recommendations for the future.

Key words: history, translation, machine translation, neural translation, history of machine translation, future of machine translation, Google Translate, SYSTRAN

Table of content

List of Abbreviations

APAC ……Automatic Translation from Czech to Slovak
CAT……Computer-Aided Translation
CBMT……..Corpus-Based Machine Translation
COMECON…….Commonwealth of Independent States
EBMT……Example-Based Machine Translation
FAMT…….Fully-Automated Machine Translation
HAMT…….Human-Aided Machine Translation
MAHT…….Machine-Aided Human Translation
RBMT……..Rule-Based Machine Translation
SMT……...…….Statistical Machine Translation
NMT……Neural Machine Translation
RNN……...Recurrent Neural Network

List of figures

Introduction

The need for a meaningful communication between people existed since the time they first appeared on Earth. Currently is this fact important even more. Although we live in a multicultural society and the boundaries between nations are becoming minor, every day we meet people with whom we cannot communicate in a proper way. The fact is that we will never know all the languages that people are communicating currently. That is why we often seek help with an intermediary who will enable us to communicate with such a person. This person is a translator or an interpreter.

Translation is one of the oldest professions in the world. In the past, exclusively people did the translation of texts, but this has thanks to the inventions of various technologies changed over time. Translators have been constantly trying to speed up the translation process because it is not only time-consuming but expensive, as well. In the field of Machine Translation, we find its supporters as well as those who absolutely reject this tool.

The Machine Translation as a subject of study has been part of Translatology for decades. Together with the development of the idea of automatization of the translation were also discovered new theories and possibilities of creating programs, which would make this automatization of the translation possible. With advances in the field of technology, the number of Machine Translation programs has increased as well.

At present, there is a growing demand to translate texts from one language into another. The easiest way to translate is to entrust a translation of the text to an experienced translator, but this process is more time demanding and expensive. The reason why are automated Machine Translators getting into the forefront is their speed and price. On the other hand, the translation is not as good as the translation by experienced translators. Because of lower quality, the Machine Translation is still followed by the Post-editing process. As the development of information technology is spreading at an incredible speed, nowadays an increasing number of translators are using a machine translator for their translation. The main purpose of the Machine Translation is to quickly and cheaply translate text from the source language into the target language using translation tools. Among the most popular Machine Translation tools today is Google Translate, one of the most popular translators, and is freely available to the public on the Internet.

In this paper, we will elaborate on the concept of Machine Translation, its history and categorisation. We will highlight the advantages and disadvantages of the Machine Translation and we will show it in some examples as well. Later, we will introduce a Neural Translation and two modern translation tools (Google Translate and SYSTRAN). We will explain how they work and evaluate their use today.

In the practical part, we will evaluate the results and analyse the questionnaire focused on the extent of the use of the Machine Translation by freelance translators and translation agencies in Slovakia. Later, we will try to specify the trends in the development of the Machine Translation in Slovakia.

1. The current situation of the topic at home and abroad

Recently, the Machine Translation is coming to the forefront in the translation industry. The name has its origin in the English word Machine Translation often abbreviated to MT. So far, it is unclear about its usability in the Slovak market. What matters here is how to face the IT evolution in the field of translation. Is the expansion of this industry related to Slovakia as well?

Professional translation is done in special programs, so-called CAT tools. We call this translation a Computer-Aided Translation. These programs help the translator to produce better and more efficient translations. Sometimes people who are not involved in the translation industry mistakenly think that Computer-Aided Translation and the Machine Translation are the same. It is not like that. It is a completely different translation process. The Machine Translation is solely the result of a computer program and the translator does not affect the translation process.

The first Machine Translations began in the 1950s, when an automatic translator was created, which translated sentences from Russian into English. Based on the Georgetown experiment, several methods of translation have evolved over time. These translations use different methods or techniques. On the present, there is a relatively large number of the Machine Translation methods. The most frequently used are as follows:

- Rule-Based Machine Translation (RBMT)
- Statistical Machine Translation (SMT)
- Hybrid Machine Translation (HMT)

The Hybrid Machine Translations focus on combining the best features of different Machine Translation approaches. In most cases, the Machine Translation is not usable immediately and needs to be post-edited. The basic idea of the Machine Translation is to produce a translation in such quality, which will reduce translation time and the total cost of translation as well. On the present, it is very popular to implement linguistic functions in the Statistical Machine Translation systems, or to modify the basic architecture of Rule-Based Machine Translations by adding data regarding statistical knowledge. Moreover, there are other ways of hybridising Machine Translation, which consist of a combination of a wide variety of translation paradigms. The Machine Translation is currently in the centre of interest and the best technologies have been able to deliver results. However, overall progress has slowed over time. This does not necessary mean that technologies such as

Statistical Machine Translation are useless, but the time when we have achieved beside high productivity, also quality is outdated. Further progress will require more and more effort.

In the translation market, we can divide users of the Machine Translation into translation agencies and freelancers. Freelancers use the Machine Translation to help them translate on a regular basis. The best-known example of this is the Google Translate and Bing Translator. These services are available to the public free, but both are limited for professional use. While Microsoft offers 2 million words a month free, Google sells its services from the very first letter. The risk is the disclosure of the translated content to the service provider, which saves it for further public use. In general, the output quality is too low for professional use.

Agencies are for their clients interested in more sophisticated and customised solutions. These solutions represent a lengthy system setup that requires linguists and terminologists who are experts in that particular language combination. The advantage of this approach is the higher quality of the translated text. The disadvantage is the need for a large number of already translated text. Despite all the efforts made, the result may not meet expectations. The Slovak language contains many linguistic rules that can be solved by Rule-Based Machine Translation, but there are even more linguistic exceptions. Moreover, the Slovak market is together with its needs so small that it is almost impossible to find a subject that produces standardised texts at least in hundreds of thousands of words per year.

1.1. Machine Translation

The Machine Translation was throughout the centuries the main topic for many researchers. Some of them thought that the Machine Translation will easily replace a human factor and there will be no need for intervention in the future. “Machine Translation will displace only those humans who translate like machines. Humans will focus on tasks that require intelligence.” (Arle Richard Lommel, 2016)

With the development of computers, there were strong efforts to reach the point of full mechanization of translation. In the last decades, we can see massive improvements thanks to the AI (Artificial Intelligence). “The mechanization of translation has been one of humanity’s oldest dreams. In the twentieth century, it has become a reality, in the form of computer programs capable of translating a wide variety of texts from one natural language into another. But, as ever, reality is not perfect. There are no translating machines that, at the

touch of a few buttons, can take any text in any language and produce a perfect translation in any other language without human intervention or assistance. That is an ideal for the distant future, if it is even achievable in principle, which many doubt." (Hutchins, 1992, p. 1)

1.1.1. Characteristics of Machine Translation

The term Machine Translation represents various computer systems that translate text from the original language into a target language without the help or with the help of a translator. This term does not include computer programs that serve the translator as a translation aid – i. e. text editors, electronic dictionaries, terminology databases, automatic spell checkers etc.

This concept includes systems whereby the translator or other user helps the machine to create a target text in the form of various online interactions or processed text editing. This type of Machine Translation belongs into the "Computer-Aided Translation" group divided into two basic subgroups:

- A person does Machine-Aided Human Translation – the translation and the machine provides several tools to help with the translation
- A machine does Human-Aided Translation – the translation and the person is responsible for text editing. (Hutchins, 1992)

The basis of the Machine Translation is to fully automate this this process and to eliminate the need for human factor.

Technology has become an inseparable part of people's lives. The Internet is coming to the forefront very rapidly, and it provides a wealth of information in various languages. Even though there are people who have the opportunity to learn a few foreign languages, they are still involved in effective mediation of global social information, and it is not enough to know one or two foreign languages. There are simply too many languages and each one is important to someone. However, a computer, notebook, tablet, or mobile phone is already part of the "basic equipment" of every person. Translators have become no exception in this field and computers make it easier and more efficient since the second half of the twentieth century.

Why should we be interested in using computers for translation? The answers are simple. There is a huge amount of texts that need to be translated and the claim that the translators can cover the whole area is not justified. Another reason is that some types of texts, especially in science and technology, are for translators uninteresting, and it is easier to rely on artificial intelligence. The third reason for a massive development of the Machine Translation is the field of terminology in a translation. Entrepreneurs, especially large corporations, companies, factories, etc. require constant stability in the field of terminology. While computer memory is consistent, translators tend to constantly seek language variations, which can easily disrupt the permanence of terminology. There are also many situations where is no need to insist on the quality of translation and in such circumstances is the Machine Translation much more efficient strategy. (Hutchins, 2005)

The term Machine Translation may be unclear very often. The Machine Translation is above years of development, but it often has a number of definitions. European Association for the Machine Translation understands Machine Translation as "the application of computers to the task of translating texts from one natural language to another. One of the very earliest pursuits in computer science, the Machine Translation has proved to be an elusive goal, but today a reasonable number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful in a number of specific domains." (EAMT, 2015)

The main objective of the Machine Translation is to produce quality output in effective, quick and cheap way. In recent years, the Machine Translation has made considerable progress in speed and quality. However, this does not mean that the translators should lose their place in society in the near future. The main goal of using the Machine Translation is a quick translation of important information for wide audience. (Munková, 2013)

1.1.2. Basic categorization of Machine Translation

There are many ways how to categorize the Machine Translation. We can categorize it for example as follows:

According to the number of contained languages:

- Bilingual systems – translation between two languages
- Multilingual systems – translation between three and more languages

According to the translation direction:

- One-way translation – translation from a source language to a target language
- Two-way translation – translation in both ways in two chosen language combinations

According to the specialization:

- Non-specialized – focused on language for wide audience
- Specialized – focused on language of a specific discipline in chosen language (Hutchins, 2005)

The most reliable translation systems today are one-way bilingual systems that are focused on specialized texts, which require only minor modifications made by a translator. The main goal is to achieve a multilingual system with as many languages as possible in two-way translations for wide audience as well as for specialized disciplines. Nowadays, despite all its shortcomings, mistakes and imperfections, the Machine Translation is the cheapest, fastest and most accessible form of translation.

Two elementary types are required when categorising the Machine Translation. The first type is a fully automatic translation that attempts to translate sentences and text as a whole unit. This means that in the translation process no human intervention is required. However, the quality of the produced output is often low and the resulting product requires human intervention in the Post-editing process. On the other hand, there are translation aids that provide linguistic assistance to translators. Those are most frequently dictionaries and databases of already translated texts that made the translation process easier and faster. (Hutchins, 2005).

There are generally several levels of translation due to their automation:

1. Fully-automated Machine Translation (FAMT)

2. Human-aided Machine Translation (HAMT)

3. Machine-aided Human Translation (MAHT)

Under **Fully Automated Machine Translation** is understood a program that is capable of perfect translating between two languages without any human intervention.

The biggest advantage of the fully automated Machine Translation is the fact that it does not require a Pre-editing or Post-editing actions. This type of translation is capable of

working on its own with perfect output quality, no mistakes and are available on-demand as well.

On the other hand, the biggest disadvantage of the fully automated Machine Translation is the fact that there would be no need for professional translators because they would not be able to compete with such a technology.

Human-aided Machine Translation is a program that can do the translation process, but because of the low output quality, human intervention through Post-editing is required. The role of the editor in this method before, during and after the machine has completed the process is very important.

The advantage of these programs is that they are mostly available on the internet free. It means they are accessible for wide audience. The use of such programs is essential when we want to understand the content of a particular text.

The biggest disadvantage is the need of Pre-editing and Post-editing. In terms of quality, these programs are far behind the fully automated Machine Translation systems and Machine-aided Human Translation systems.

Machine-aided Human Translation is such a program that assists the translator in the translation process whenever the translator requires. Modern Machine-aided Human Translation programs are based on translation memories created by the translator itself. (Absolon, 2018)

Machine-aided Human Translation systems are mostly paid and requires physical intervention in order to create large text corpora, that are assigned for different text types. The output quality is much better than the Human-aided Machine Translation systems, but these programs rely on the specific text corpora and are limited to its own vocabulary.

1.2. The history of Machine Translation after the 2nd World War

Shortly after the first researches the so-called "computer machines" have also begun to be used as a support for the translation of languages. Based on a letter from Warren Weaver of the Rockefeller Foundation to Cybernetics Norbert Wiener we can dated the beginning of this field back to 1948. Two years later W. Weaver wrote a memorandum in which he put a various number of suggestions in the forefront. These suggestions were based on the knowledge of coding acquired during the war. This was also the reason for carrying out the Machine Translation researches at several US universities.

In 1954, there was the first public demonstration of the Machine Translation feasibility in collaboration with IBM and Georgetown University. Although they used only a very limited vocabulary and grammar, they made a good impression encouraging massive financing of the Machine Translation in the USA and inspired the formation of the Machine Translation projects all around the world.

The first systems consisted of extensive bilingual dictionaries that offered two or more equivalents in the target language for the words inserted into the machine. They also ensure that the final words are correctly aligned, so the target language word order will be followed. However, they found soon that the word-order rules of these machines were too complex. Numerous projects have been based on the recent achievements of modern linguistics, particularly those of formal grammar models, which have provided a big hope for translation improvement.

Optimism was particularly noticeable for the first decade of research. Researchers have assumed that advances in the Machine Translation will be quick, applicable, and their modernity will stun the world. However, these assumptions were quickly confuted due to semantic barriers, because their solutions were still far away. At this time, several translation operating systems already existed. One of them was for example the Mark II developed by IBM and Washington University, whose output quality was unfortunately not gratifying. (Somers, 1992).

In 1964, the US government set up a special commission – ALPAC – (Automatic Language Processing Advisory Committee) and entrusted it to investigate the situation of the slow Machine Translation research. Two years later, the Commission announced that the Machine Translation is slower, less accurate, and more expensive than the translation made by the translator and therefore there is no perspective of the usefulness of the Machine Translation in the future.

After all these events, funding of these projects was greatly reduced and researchers turned their attention to translation aid tools, especially automated dictionaries.

The results of the ALPAC investigation have had a disfavourable impact on the Machine Translation in many countries around the world and there has been a period of great disappointment. Research and funding have continued exclusively only in Germany, France and Canada, but it was more successful.

The first two countries are linked to the creation of SYSTRAN in 1969, which is being developed also today and is used as a basis by Internet giants such as Yahoo! and

Google. In 1976, a METEO system was launched in Canada, which originally served to translate weather forecasts for the Canadian region from English into French. Currently, this system handles about 20 to 30 thousand weather forecasts per day.

Several countries were inspired by the previous Machine Translation successes and turned their attention to it again. In addition to the morphological and syntactic analysis, the semantic analysis and analysis of non-linguistic elements was added to some extent as well.

In 1990s, there was a breakthrough event a shift away from the Machine Translation research towards the development of practical tools to help translators perform their profession and develop multilingual translation systems. Translation devices used no longer sets of semantic and syntactic rules but the so-called corpora text access.

The Machine Translation is mainly focused on a fast-available translation for personal computer users but the modern tools for translators continue to develop as well. The most important aspect is the speed of the translation process from the beginning of translation to the output text. Improving the quality of the output text bear back to the background.

The Machine Translation begins to be used at a professional level to quickly and efficiently translate documentation for various products into the new markets. The demand for immediate translations into new languages is constantly rising. Extensively popular are the translation software and automated translation services available online such as Yahoo! and Google Translator. (Horwood, 1986)

1.2.1. Machine Translation in Czechoslovakia

The beginnings of Machine Translation research in Czechoslovakia date back to the 1950s when the connection between the language and computer technology began to be debated for the first time. This research was chaired by Peter Sgalla and was conducted mainly at The Charles University in Prague. At the end of the 1950s, first people interested in the Machine Translation in Czechoslovakia were linguists from the Faculty of mathematics and physics at The Charles University and several teachers and students from the Faculty of Philosophy at Charles University as well, who were particularly interested in the theoretical level of linguistics.

Another series of experiments focused on texts from the field of electronics and computer technology was run on the improved computer system EPOS I in 1966. For this

occasion, an electronic text corpus consisting of about one hundred thousand words from English to Czech was compiled. These modest steps were the beginning of the journey towards a high-quality Machine Translation system. In spite of the American optimism that prevailed in the world at this time, Czechoslovak researchers retained a realistic view of this problem. They did not expect that it would be possible to create a system that perfectly translates the text from the source into the target language in the near future. Therefore, we can talk about the absence of optimism and the period of disillusion.

From 1966 to 1976, the Machine Translation research in Czechoslovakia stayed still and all experiments were aborted. In 1976 was created a new project called APAC ("Automatický preklad z češtiny do angličtiny" – Automatic translation from English to Czech). The author of this experiment was Zdeněk Kirschner. The first of the series of efforts took place in cooperation with the TAUM group of the University of Montreal in 1978. This project was replaced by an improved APAC-2 in 1982, which was experimental until the end of 1988 and later inspired the system RUSLAN in translations from Czech to Russian.

Secondly, the Machine Translation project focused on translating Czech manuals into Russian started in 1985. These experiments were active at the Charles University until 1990. The aims of this project were both theoretical and practical. From the theoretical point of view, they wanted to develop a better background for the Machine Translation between Czech and closely related familiar languages. From the practical point of view, on the other hand, this experiment should serve as a basis for export purposes of COMECON countries (Commonwealth of independent states) and translate from 500 up to 1000 pages a year. (Hajič, 1992)

For those days, RUSLAN translation system was a very complex and complicated computer program. "RUSLAN is a unidirectional bilingual system. The translation scheme is transferlike in the sense that no intermediate pivot language is used. However, many simplifications were made with the expectation that the close relationship between Czech and Russian does not require a full-fledged transfer scheme. The result of this is that the system resembles in some respects the so-called direct method." (https://www.erudit.org/fr/revues/meta/1992-v37-n4-meta338/002996ar.pdf, 2019-01-06)

2. The process and methods of Machine Translation

In this chapter, we will speak about the process and methods of Machine Translation. Moreover, I will describe different approach, which are affective in current research and were affective in the past. Rule-Based, Dictionary-Based, Transfer-Based, Interlingua-Based, Corpora-Based, Example-Based and Statistical-Based Machine Translations will be discussed in the following chapters.

We can say that there are two main approaches in the field of Machine Translation:

- **Rule-Based Machine Translation** (RBMT)
- **Corpora-Based Machine Translation** (CBMT)

As mentioned before, the Machine Translation systems are designed to follow certain criteria, which are known as so-called "approaches". Linguistics architecture is closely related to Rule-Based úand computational linguistics, on the other hand, to Corpora-Based Machine Translation. (Munková, 2017)

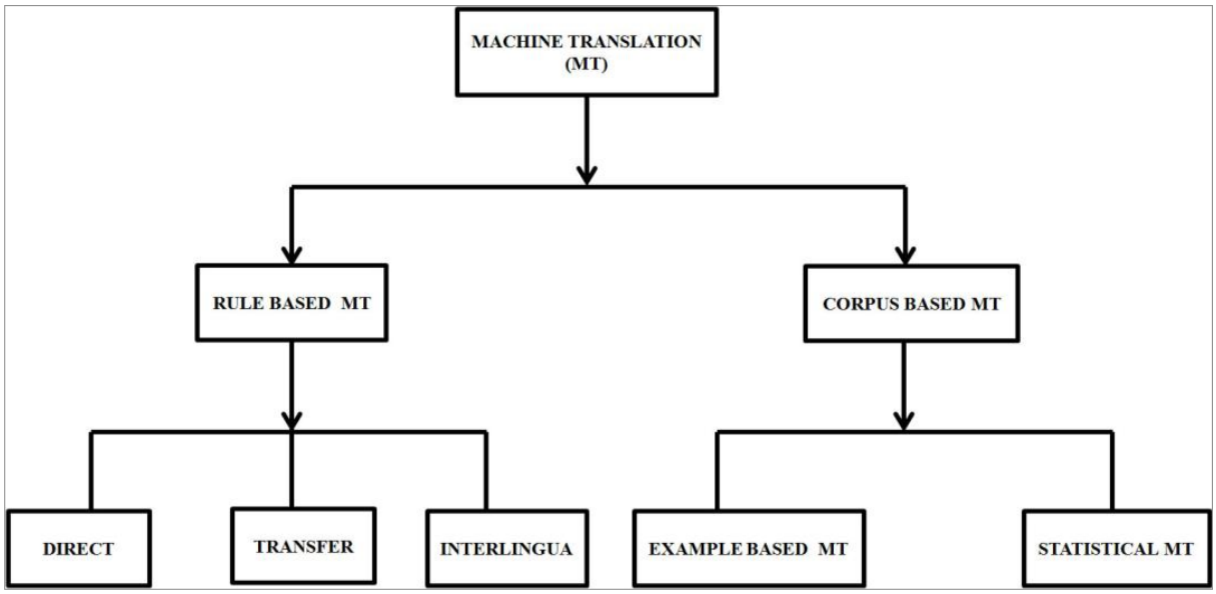


Fig. 1 – Types of Machine Translation (Anju, 2014, p. 19)

2.1. Rule-Based Machine Translation

Rule-Based Machine Translation is an approach requiring wide range of grammatical rules and is built on linguistic analysis. This approach requires analysis, substitution of source text and synthesis of the source text in the target language. The most important aspect

of Rule-Based Machine Translation is the creation of clear and explicit lexical and structural text representation.

2.1.1. Dictionary-Based Machine Translation

The first level represents direct and lexical oriented way. Source text is morphologically analysed. Word forms and suffixes are simplified to their basics. Systematically, the Machine Translation system looks for word for word equivalents in the target language. This approach is very complex and requires enormous bilingual dictionaries.

Program generates a translation from the source to the target language and does not take into the consideration the morphological analysis, words or context. The Dictionary-Based Machine Translation is essential for translation of longer phrases. (Sindhu, 2017)

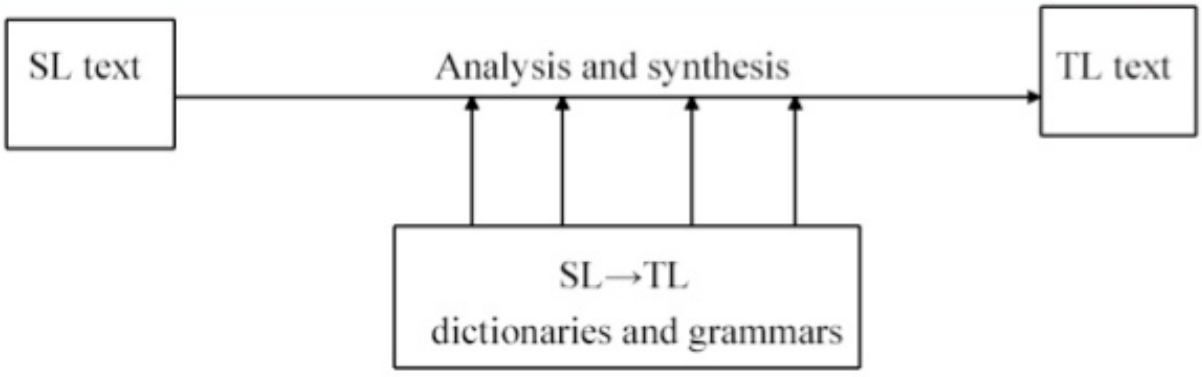


Fig. 2 – Dictionary-Based Machine Translation (Sindhu, 2017, p. 20)

2.1.2. Transfer-Based Machine Translation

The second level represents the transfer, which executes linguistic analysis on three different levels. (Munková, 2017)

- **Text analysis**

In Machine Translation, the machine primarily deals with linguistic analysis of the source text. While using pre-defined rules machine perceives the text as a whole unit.

a. Syntactic analysis

At this stage of the analysis, the machine looks at the syntactic level of the sentences. The machine examines individual components, their placement in relation to other aspects and creates an idea of the separate structure of different sentences. An important part of this phase is to leave all parts in comprehensible form for the machine.

b. Morphological analysis

In the next step, the machine makes a morphological analysis of units that were created during the syntactical analysis. Bilingual dictionaries contain only the basic form of each word.

c. Semantic and Contextual Analysis

This phase is critical for a correct understanding of the translated text. The machine analyses relationships related to context and meaning. While he is unable to understand these relationships, the minimal pre-defined rules will modify what phenomena he should look at and how to analyse them. (Chéragui, 2012)

- **Text transfer**

a. Transitional form of text in the original language
b. Transfer to a transitional form of text in the target language

After the end of the analysis phase, the machine will compile a "temporary idea" of all collected data on the original language. The task of the transfer phase is to modify this idea into a form suitable for subsequent translation into the target language. (Chéragui, 2012)

- **Text synthesis**

a. Semantic and contextual synthesis
b. Morphological synthesis
c. Syntactic synthesis

The final phase runs counter to the initial phase and it is opposite to the text analysis. Here is a synthesis of the output text in the target language. (Chéragui, 2012)

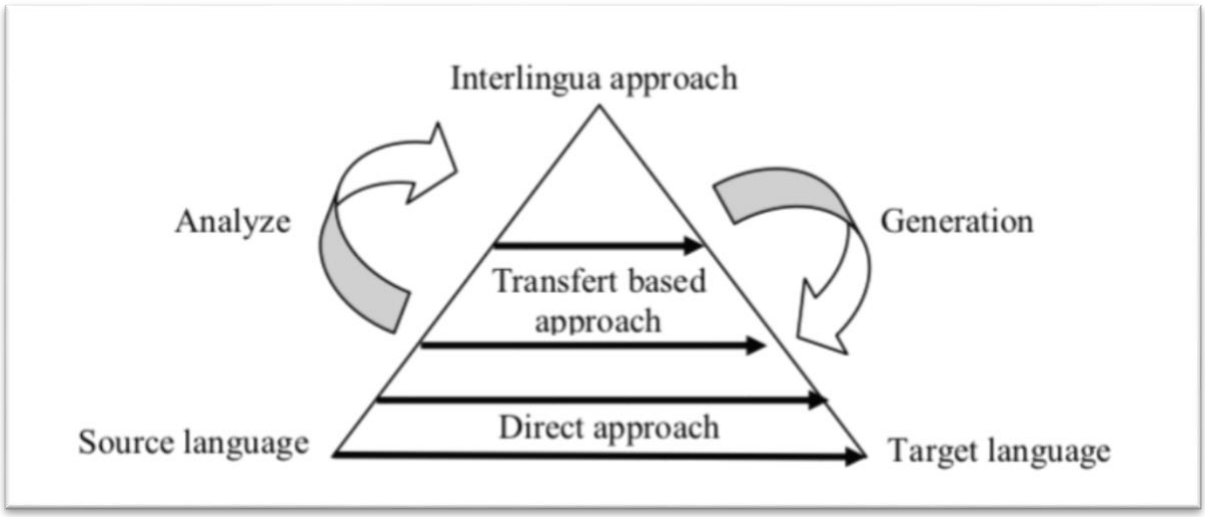


Fig. 3 – The Vauquois triangle (Vauquios, 1968)

2.1.3. Interlingua-Based Translation

The next type of the Machine Translation is an Interlingua-Based Machine Translation. It is similar to the Transfer-Based Machine Translation but there is one fundamental difference between them. In the Interlingua-Based Machine Translation is the text is not translated directly from the source to the target language but from the source to the Interlingua language and then to the target language. Interlingua is an intermediary element in which are added source and target languages.

As far as the meaning of the text is concerned, Interlingua-Based Machine Translation focuses on a more in-depth analysis of the language. It represents the meaning of the text independently of the particular language. In this method is no transfer and a large number of components are not required to create language pairs. It is a more sophisticated approach that includes bi-directional grammar for both languages and their analysis as well.

It consists of two levels:

- from an **analyser**
- from a **generator**

Analyser transforms the source text into interlingua and the generator transforms interlingua into a target language. (Munková, 2017)

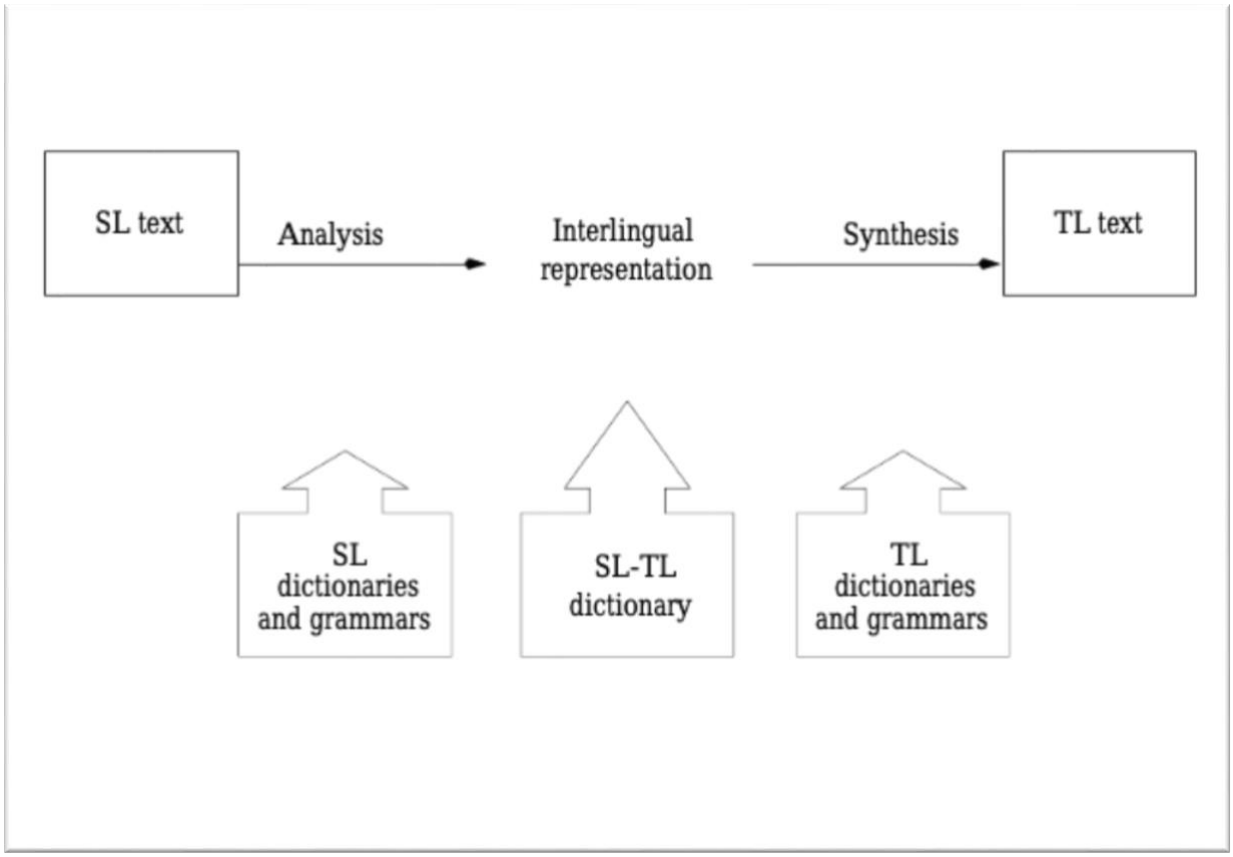


Fig. 4 – Interlingua-Based Machine Translation (Orhan, 1998, p. 10)

2.1.4. Corpus-Based Machine Translation

Corpus-Based Machine Translation (CBMT) does not include rules and grammar but an aligned bilingual corpus. The CBMT system includes corpora in a machine-readable and these corpora are written or spoken. For the Corpus-Based Machine Translation is also required translational knowledge from bilingual corpora. Corpus-Based Machine Translation systems are applied in the Example-Based Machine Translation (EBMT) and Statistical Machine Translation (SMT). (Munková, 2017)

"The corpus-based approach is also termed as an empirical approach in which the knowledge sources to develop a Machine Translation system are computed automatically by analysing example translations. A major benefit of empirical approaches is that Machine Translation systems for new language pairs and domains can be developed rapidly and with ease, provided sufficient training data is available." (Shah, 2012, p. 13) Figure 2 shows the architecture of a Corpus-Based Machine Translation system.

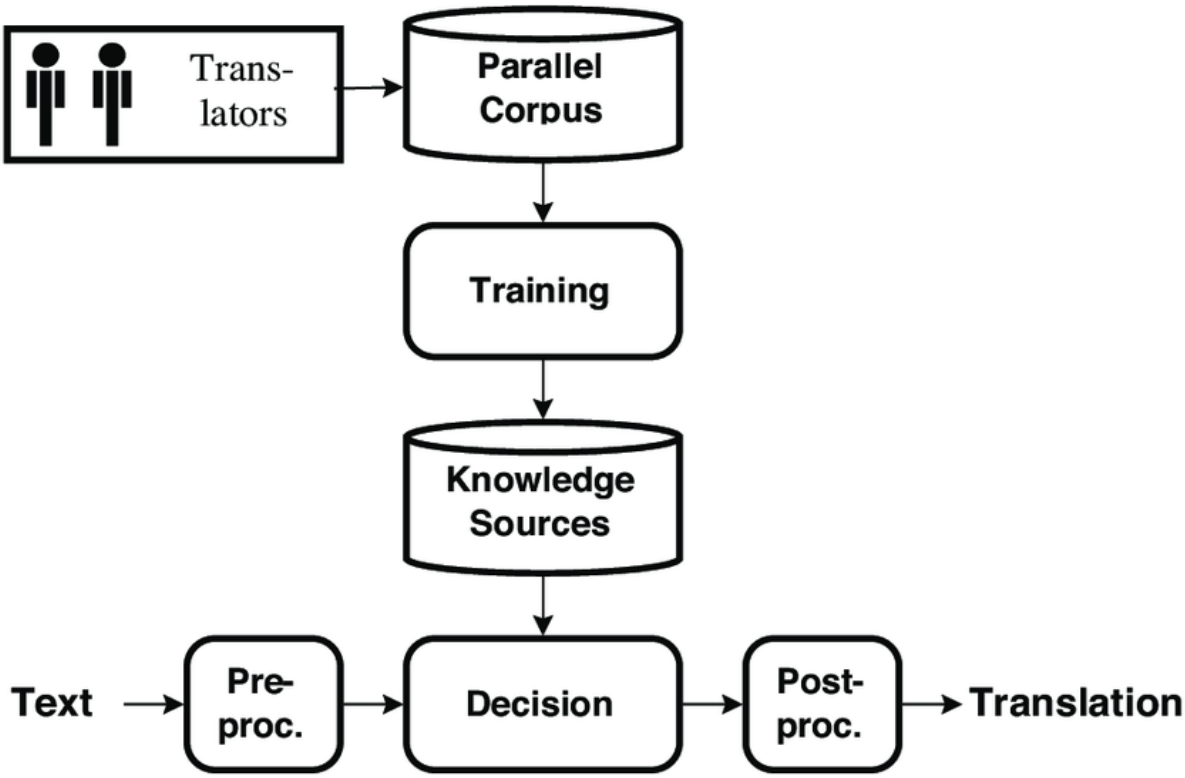


Fig. 5 – Corpus-Based Machine Translation (Och, 2002)

2.1.5. Example-Based Machine Translation

Example-Based Machine Translation uses existing translations as the main source of its Machine Translation. It creates new translations using a previous translated examples and sentences, so the Example-Based Machine Translation is limited to its bilingual corpora. These texts contain repetitive translations of sentences that can be reused in full form or can be divided into several fragments and later combined in various ways and used in new translations. The translation contains three stages:

- matching, finding identical segments in a parallel corpus
- aligning, identifying the corresponding segments
- combining the corresponding segments in an adequate manner (Munková, 2017)

This method is following the opinion of the professor of linguistics Makot Nagao who stated that the translator does not make an in-depth linguistic analysis when translating

simple sentences. Instead, the translator divides the sentence into smaller separate units and produces output text within the synthesis. The machine works very similar to this principle.

In addition to word and translation, these dictionaries contain use in various composites and idioms, which increases their efficiency in word translation. In spite of the use of prepared bilingual corpora from translators, the example method represents only little progress in Machine Translation. (Bhattacharyya, 2015)

"In the example-based approach, translation is performed by analogy. The source sentence to be translated is searched in parallel corpora, translation examples are extracted and combined to generate the target sentence." (Shah, 2012, p. 13)

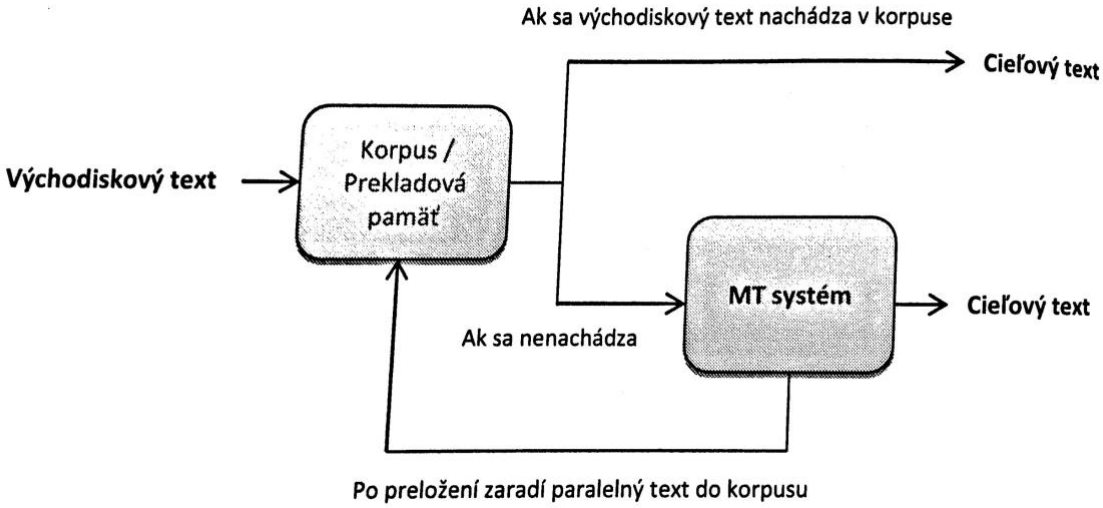


Fig. 6 – Rule-Based Machine Translation (Munková, 2017, p. 20)

2.1.6. Statistical-Based Machine Translation

"Modern machines can be programmed to learn from empirical data with the de- sign and development of sophisticated algorithms. The algorithms automatically extract knowledge to recognize complex patterns and make intelligent guesses based on data. This paradigm is named machine learning and is widely used to solve many problems. Statistical-Based Machine Translation is based on translation by learning statistical models. These models are learned by already available translated text." (Shah, 2012, p. 14)

In this type, the target text is generated based on statistical methods and probabilities. It deals with bilingual corpora in which there is a huge database of bilingual texts. The Statistical-Based Machine Translation uses two translation approaches. One is based on words and the other on phrases or sentences. The good example of this translation approach is Google Translate.

The Statistical-Based Machine Translation is one of the most up-to-date methods. The translation uses the Statistical-Based Machine Translation method and applies it to bilingual text corpora. For proper and efficient operation, the machine does not need to have any defined linguistic knowledge. The output text quality depends on the size of the available bilingual corpus.

The basis of this method is that it attempts to simulate a natural way of learning language and creates very simple and basic kind of artificial intelligence. By the analogy from correctly translated bilingual corpora, it selects the most appropriate formulation and translation.

Statistical-Based Machine Translation is similar to Example-Based Machine Translation. Both approaches require extensive bilingual systems, but the difference between them is that Example-Based Machine Translation systems are learning based on examples and Statistical-Based Machine Translation systems are learning based on statistic. The Statistical-Based Machine Translation consists of two parts:

- obtaining both translational model and a language model of the target text
- encryption of the source sentence

The translational model characterising the adequacy of the translation serves us the possibility of the probability that the word or phrase is translated correctly regarding the bilingual corpus.

The language model is characterised by the consecutiveness of the translation providing us with information about the probability that the word or phrase is well structured and trained on the monolingual corpus of the target language.

The decoder takes the source sentence and all available information including the language model, translation model and the translation probability and generates the target sentence. The advantage of Statistical-Based Machine Translation is the self-learning aspect. The disadvantage is the training itself and the translation equivalence. One word could be assigned to more words and vice versa. Various idioms and words without the equivalence in the second language also exist. (Munková, 2017)

2.1.7. Hybrid-Based Machine Translation

A relatively new among all the above-mentioned methods is the Hybrid-Based Machine Translation that combine the elements of multiple methods in their translation systems (most often two). This fragmentation is characterised with the fact that the different phases of the translation process are solved by various methods. (Munková, 2017)

"Hybrid Machine Translation (HMT) integrates the strengths of rationalism method and empiricist method. Essential of HMT is to integrate the core of MT engines. Multiple-engine HMT integrates all available MT methods, applying to their benefits, in order to improve qualities of output. HMT can also be constructed based on RBMT. Translation process is completed by coupling with SMT or EBMT that is employed to solve problems with certain parts. According to present requirements, the most popular combinations comprise "rule-based MT vs. statistical-based MT", "rule-based MT vs. example-based MT", and multiple combinations." (Xuan, 2012, p. 3018)

From our point of view, Hybrid-Based Machine Translation systems are still the best translation systems. These programs are taking the best from different translation approaches and the result is the enhanced output text quality.

Hybrid-Based Machine Translation combines mostly Statistical-Based Machine Translation and Rule-Based Machine Translation. This translations system is most essential when translating between language pairs with low frequency of the translation. This is caused because of lack of language data for self-learning of a particular program. Texts translated with Hybrid-Machine Translation systems are still containing mistakes but the output text is understandable and correct.

3. Machine Translation tools

If you decide to use a Machine Translation, you need to consider whether it is actually suitable for it. Those special texts in which this method is valid include official correspondence, template documentation, technical or scientific documents with limited vocabulary, as well as documents written in the artificial (controlled) language. Controlled languages are usually derived from a native language, in which is minimised grammar and eliminated ambiguity.

These languages include simple English, Esperanto or Interlingua. The least suitable for the Machine Translation are literary works because they require large punctuation quality and there is a strong emphasis on the context and the authors' intention.

3.1. Google Translate

On the present, the Google Translate is the most advanced example of the Machine Translation based on the Statistical-Based Machine Translation approach. The engine of the entire system are enormous bilingual and multilingual text corpora, which have millions or even billions of words. Prominent languages mean mostly official languages of the European Union, as well as the official languages of the United Nations Organisation.

Until 2016, the Google Translator used the Statistical-Based Machine Translation approach, which does not require source or target language knowledge from its developers. Therefore, it is not surprising that most of the Google Translate developers speak only English. Because they do not have to deal with lexical and grammar issues, they can concentrate on developing the translation system and implementing new bilingual documents.

The translation process of Google Translate is relatively simple. The second chapter introduced the Statistical-Based Machine Translation method and based on that Google Translation software does not need to have any language skills and uses the natural way of learning a language. Compared to other methods, it does not need the analysis and synthesis, grammar or vocabulary. Instead of these steps characteristic for previous and older translation programs, Google Translator looks for a similarly and previously translated equivalent which is then applied in a translation. The whole process is extremely fast and the translated phrase is not the first result that the translator finds. It compares tens, hundreds,

and sometimes-even thousands of results and selects the one that is most adequate for the context. (TAUS, 2016)

In 2016, Google Translate introduced a translation based on an artificially created network of texts, where the system considers as a basic unit the sentence and not individual words. This method is also called Neural Machine Translation. Macduff Hughes, the Director of Engineering for Google Translate, explains that the translation system compares the document that is very similar to the source text and therefore could be a possible source for the translation. This theory is based on that a large amount of unverified data is more than a small amount of verified data.

While Google, for example, cooperates with the European Patent Office and other institutions that provide them with a great amount of data, there is only a little chance that they will have some weight because their servers work with one trillion data. Today they offer translations into more than 103 languages and translate more than 100 billion words a day. Interestingly, more than 3.5 million people have already contributed to improving this program with the so-called Google Translate Community. The country where is Google Translate used most frequently is Brazil. (Turovsky, 2016)

3.1.1. Google Translate from our point of view

It might be claimed that the Google Translate (and the Statistical-Based Machine Translation in general) is not actually a translator, because it works by searching for existing phrases in the target language rather than translating it. It works on the principle of a search engine, wherewith is Google Inc. known for. This approach makes it easy for the system to add other languages and language combinations without the need for qualified linguistic staff.

A great advantage is its speed, which is incomparable with a translator (a person). On the other hand, Google Translator is dependent on a translator. This dependence on human translation means the constant need for live translations. If the machine is about to translates a phrase that does not exist in an available binary or multilingual corpus, it will most probably fail.

Another advantage of the Google Translate is extensive and expanding language availability. While many combinations do not have sufficiently extensive bilingual corpora,

the translator can get through this thanks to the so-called intermediate language. However, this process is rare especially in newly added languages in which this function would be very useful.

This method also has many deficiencies in obsolescence of vocabulary and in grammatical changes in terms of natural language development. We can assume that if it comes from an existing corpus that contains the vast majority of older documents, it prefers an older expression. This deficiency could be easily removed by updating corpora, which is practically impossible, when we look at the size of the corpus.

From a technical point of view, the operation of such a massive system is financially very difficult. It requires high-performance technology to process complicated search and analysing processes in bilingual corpora, as well as large storage capabilities and access to millions of words.

3.1.2. Google's Neural Machine Translation approach

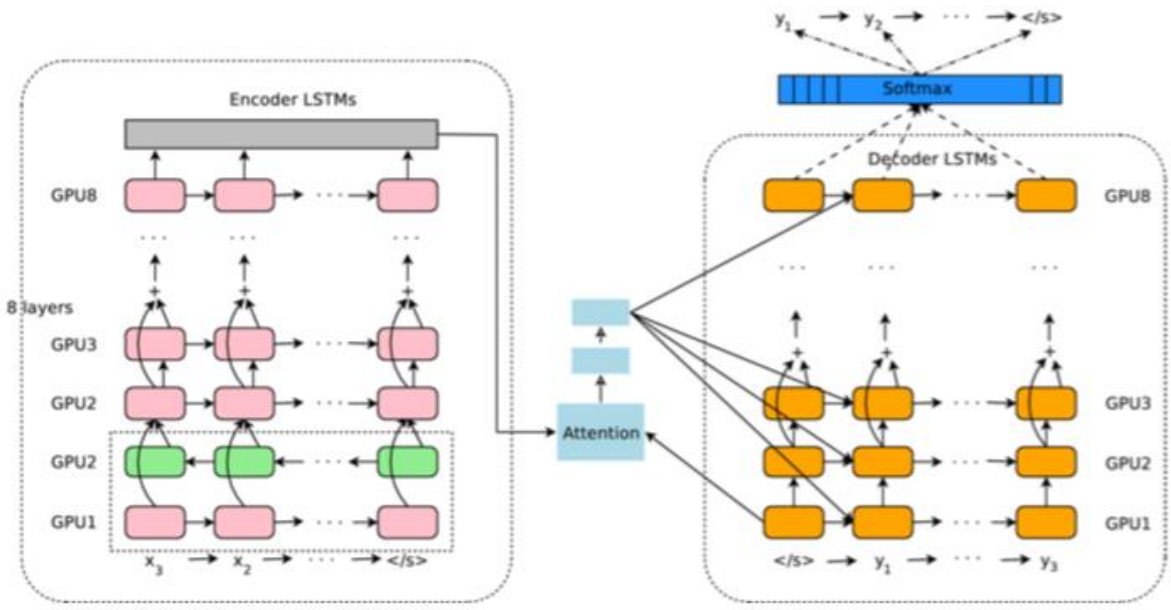


Fig. 7 – Google's Neural Machine Translation approach
(https://arxiv.org/abs/1609.08144v1, 2019-01-09)

In 2016, Google introduced its own Neural Machine Translation system. It is a very complex and complicated process. "The model architecture of GNMT, Google's Neural Machine Translation system. On the left is the encoder network, on the right is the decoder

network, in the middle is the attention module. The bottom encoder layer is bi-directional: the pink nodes gather information from left to right while the green nodes gather information from right to left. The other layers of the encoder are uni-directional. Residual connections start from the layer third from the bottom in the encoder and decoder. The model is partitioned into multiple GPUs to speed up training. In our setup, we have 8 encoder LSTM layers (1 bi-directional layer and 7 uni-directional layers), and 8 decoder layers. With this setting, one model replica is partitioned 8-ways and is placed on 8 different GPUs typically belonging to one host machine. During training, the bottom bi-directional encoder layers compute in parallel first. Once both finish, the uni-directional encoder layers can start computing, each on a separate GPU. To retain as much parallelism as possible during running the decoder layers, we use the bottom decoder layer output only for obtaining recurrent attention context, which is sent directly to all the remaining decoder layers.

The softmax layer is also partitioned and placed on multiple GPUs. Depending on the output vocabulary size we either have them run on the same GPUs as the encoder and decoder networks, or have them run on a separate set of dedicated GPUs."
(https://arxiv.org/abs/1609.08144v1, 2019-01-09)

3.2. SYSTRAN

SYSTRAN is one of the oldest and one of the few translators that are still developing also today. The age of this translation system is also evidenced by the fact that the translation process used to use the Dictionary-Based Machine Translation approach, which is now out of date. Today, it uses a more efficient method of Machine Translation, which is called a Hybrid-Based Machine Translation. This means that compared to the Google Translate (Statistical-Based Machine Translation), SYSTRAN is a combination of both Rule-Based Machine Translation and Statistical-Based Machine Translation approach. Since this translator is also based on predetermined linguistic rules, the linguistic aspect is incomparably more advanced.

The quality of translations has improved significantly since the birth of the translator and many European Union institutions and external users have begun to use it afterwards. SYSTRAN has become an indispensable tool in many professions because of its reliability, ease access and a simple user interface. However, it is a major disadvantage for Slovak-speaking users that they are not available in Slovak but only in the following languages:

English, Arabic, Chinese, Danish, French, Greek, Hindi, Dutch, Croatian, Japanese, Korean, Norwegian, German, Persian, Polish, Portuguese, Russian, Serbian, Spanish, Swedish, Ukrainian and Urdu.

SYSTRAN has been developing most of the time especially for language combinations where English or French is the target language. Up to 85% of translations are in one of these two languages. The third is German with 8% and other languages with the remaining 7%. The target languages overcome English and French with approximately 26% each and the third is Spanish. (Chesterman – Wagner, 2014)

3.2.1. SYSTRAN from our point of view

If we look at the translation process of SYSTRAN, we see that it is significantly more sophisticated and linguistically more advanced than the Google Translate. To some extent, it is because SYSTRAN is a hybrid translator and uses the functional protocols of the Statistical-Based Machine Translation and Rule-Based Machine Translation method. It uses a statistical search in defined glossaries, but the output text affects largely the quality and complexity of the linguistic rules defined for different language combinations.

However, the complexity of SYSTRAN directly means a financial difficulty in developing such a translation system and a long time to develop a new language combination, which needs to be defined by a number of rules and to compile dictionaries to ensure the correct functioning of the translator. In this field, Google Translate has a great advantage because it is enough for this system to provide as many bilingual texts as possible and let it work with them. The disadvantage of Google Translate is that it does not translate but it searches for previously translated materials. However, we can say that both translators are dependent on human work, and they are still far from being automated.

SYSTRAN does not perceive the text as a whole, but it divides it into segments, which reduces quality and limits the translation process. If it analyses well and understands one sentence, it does not guarantee the correct translation of the next one. Language development (grammar, vocabulary) means for SYSTRAN no problem because it requires only updating the rules in the memory of the translator or changing the vocabulary elements in the dictionary. Compared to Google Translate, SYSTRAN cannot do this by itself. These changes require human intervention but are valid immediately after an update and that is very fast. The Google Translate is dependent on corpora. This means that as long as the

amount of use of the new shape does not exceed the amount of use of the old one, it always prefers the old form.

From a technical point of view, SYSTRAN with its sophisticated translation approach is less demanding when performing different processes compared to Google Translate. Predefined spelling and precise vocabulary eliminate the need to search in a large number of bilingual texts and allow it to run on less powerful machines.

4. Neural Machine Translation

The youngest approach to the Machine Translation is the Neural Machine Translation (NMT) introduced in 2013. As we have already noted in 2nd chapter and we would take into consideration a traditional types of the Machine Translation, it would be possible to include the Neural Machine Translation in the Corpus-Based Machine Translation. For self-developing, it needs parallel data and uses advanced methods in the field of statistics and machine learning. We could strictly say, Neural Machine Translation would be a subtype of Statistical-Based Machine Translation.

If we look for the simplest answer to the question of what the Neural Machine Translation is, it might sound like this: "Neural Machine Translation is the approach of modelling the entire Machine Translation process through a large artificial neural network." (Manning - Socher, 2017) We could say that the neural network is a programmed model that simulates the biological structures of neurones in living organisms and is able to learn through so-called deep learning from available data.

Most of the Neural Machine Translators are based on so-called encoder-decoder model that uses the Recurrent Neural Network (RNN) to be able to take up the sequence of the default language and predict the sequence in the target language. The source sentence is first encoded into a vector with a fixed length. During decoding, a translation from the encoded vector is generated (Bahdanau - Cho - Bengio, 2014).

Although this is a relatively new approach, experts agree that it delivers promising results. However, to make the Neural Machine Translation work properly, it must have a large parallel corpus. If it does not have enough data for self-developing, the quality gets worse. The absence of sufficiently large parallel corpora is a problem that many language pairs have to deal with and it is not just small languages like Basque, but language pairs of some major languages, such as the Russian and German combination. (Artetxe - Cho, 2017)

There are some next further challenges that the Neural Machine Translation will have to face in the future. The Neural Machine Translation has a worse quality when translating very long sentences and problems can also occur when translating less frequently used words. Regarding the advantages of Neuronal Machine Translation, the biggest to name is a

greater fluency of the generated translation or the use of a wider context for higher accuracy. (Manning - Socher, 2017)

4.1. Neural Machine Translation and Statistical-Based Machine Translation

For a better imagination, we could compare the Neural Machine Translation with the Statistical-Based Machine Translation. We find out that Statistical-Based Machine Translation works with many small subcomponents and the Neural Machine Translation tries to create and develop a large unified neural network that is capable of understanding the whole sentence and generating its correct translation. This approach attempts to analyse the sentence as a whole and then translate it as a whole, as well. (Bahdanau et al., 2014)

5. Disadvantages and advantages of Machine Translation

In this chapter we will speak about the advantages and disadvantages of the Machine Translation.

We will emphasize the disadvantages on several examples. More specifically, we will demonstrate these disadvantages with examples based on common communication, context impact, polysemy, synonyms, homonyms, idioms and dynamic language. The source Machine Translation tool, we decided to use Google Translate.

Later, we will bring up some advantages of the Machine Translation based on facts connected to the Machine Translation tools. We will also go through the 2018 Language Industry Survey and we will try to summarise this interesting survey.

5.1. Disadvantages of Machine Translation

The Machine Translation of specialized texts limited to a certain area is already achieving excellent results. Despite the constant development of Machine Translation, this form of translation is generally not reliable and cannot be compared to the translator's work. In comparison with a person, Machine Translator cannot understand the complexity of a language. In the following subchapters, we will introduce the most frequent types of the Machine Translation errors.

5.1.1. Common communication

In contrast with specialised texts that have a defined word vocabulary, common communication is one of the most problematic areas of the Machine Translation. The range of vocabulary possible used is enormous and very varied. It contains elements from all spheres of vocabulary, from standard words to slang and vulgarisms, which are often problematic for the Machine Translation and in some cases even untranslatable. The number of non-standard words is constantly growing. These words are updating their meaning and often bearing polysemy. Each of these aspects contributes to the degradation of the translating quality.

Example: *"I have a **ball**."*
Translation: *"Mám **loptu**."*
Correct translation: *"Dobre sa bavím,"*

5.1.2. Context impact

As we have suggested earlier, the largest unit with which the machine in the translation work, is a segment of a sentence or a sentence itself. In translating multiple connected sentences, the machine cannot detect it. Thus, there are considerable shifts in translation that confuse the recipient of the output text.

Example: *"Don't be scared. I'm here for you. Just **give it a shot**."*
Translation: *"Neboj sa. Som tu pre teba. Len **daj ten výstrel**."*
Correct translation: *"Neboj sa. Som tu pre teba. Len **tomu daj šancu**."*

5.1.3. Polysemy

In some cases, the translator also has the problem of finding the correct equivalent in the target language when translating the polysemantic word. It is even more difficult for the machine, even though some modern systems contain recognition algorithms for polysemantic words and can linguistically analyse them. There are only a very few mistakes in the translation because a set of machine rules defines how to deal with it. The machine cannot translate, when the individual semantic variations of the polysemantic word are the same word class.

Example: *"**The tree-top** of the tree is beautiful."*
Translation: *"**Strom** stromu je nádherný."*
Correct translation: *"**Koruna** toho stromu je nádherná."*

5.1.4. Synonyms

In synonymous terms, the machine has no tools for recognising between semantic meanings and abnormalities. In most cases, the machine sees word like a unit and considers

only its first glossary meaning for a specific word. Some more modern systems have predefined conditions for which they choose the right meaning but still have significant limitations in many ways.

Example:	*"I go to my hotel to **hit the sack** soon."*
Translation:	*"Pôjdem do môjho hotela, aby som rýchlo **zasiahol vrece**."*
Correct translation:	*"Pôjdem do môjho hotela, aby som **šiel skoro spať**."*

5.1.5. Homonyms

Homonymous are words that have the same form but different meanings - graphic (homograph) or phonetic (homophone). The Machine Translation has problems especially with homographs, as it mainly works with the graphic form of a written word.

Example:	*"His book is **on fire**!"*
Translation:	*"Jeho kniha je **v ohni**."*
Correct translation:	*"Jeho kniha je **na roztrhanie**."*

5.1.6. Idioms

Each language contains a set of custom idioms which are translated in the same way into another language only very rare. If the machine does not have a predefined translation of the idiom in the target and the source language, it works with a sentence as an ordinary idiom-free unit. Such translation is almost inaccurate and inapplicable in all cases.

Example:	*"Speak of **the devil**."*
Translation:	*"Hovorte o **diablovi**."*
Correct translation:	*"**Osoba**, o ktorej sme práve hovorili, sa objavila."*

5.1.7. Dynamic language

We have already outlined that language is constantly developing and updating. Machines that have all parameters defined by a person cannot understand these changes and

modify them in their database. Without human intervention, such databases are rapidly obsolete in the current dynamic period, dramatically affecting the quality of the output text. (Munková, 2017)

5.2. Advantages of Machine Translation

Although the source text of the Machine Translation is not always translated correctly, we can say that the Machine Translation has many advantages. One of them is its almost immediate availability. The biggest advantage of free available on-line translation tools is that they can handle large volumes of text in a relatively short time. What is necessary to point out that it is often necessary to edit such a translation.

Another advantage is that it can translate texts in different language combinations. The Machine Translation eliminates potential errors and corrects the typing errors that the translator may commit during the translation.

Finally, yet importantly, the Machine Translation works with the current level of language and grammar. The benefit is not only a translation of common words, but also their proper inflection into the correct form or a change of the word order regarding the target language. (Svoboda, 2012)

In 2018, there was a Language Industry Survey with 1285 responses from 55 countries. Regarding this survey and its results, we can see a great amount of optimism in the translation industry. Translation companies are expecting an actual growth. More precisely, "67% of the responding companies expect an increase in activity, against only 5% a decrease. This is a further improvement compared to 2017, and an obvious sign that the companies expect the actual growth in 2017 to continue." (2018 Language Industry Report, 2018, p. 8)

Regarding the 2018 Language Industry Survey, we can see a constantly growing demand for translation services. We will remember the year 2018 as a year of the Machine Translation as well "as the year in which more than 50% of both the companies and the individual language professionals reported that they are using MT in one form or another." (2018 Language Industry Report, 2018, p. 15)

Another important thing resulting from the 2018 Language Industry Survey is the fact that the market has accepted the Machine Translation. The most frequently used tool for

the Machine Translation is still the Google Translate, therefore there is no need for companies or individuals to make any investment in Machine Translation. “52% of all respondents report that they are using the site, but we see a clear difference between the various categories of respondents. While more than 70% of the respondents in training institutes report that they are using the site, only 49% of the translation companies and 52% of the individual translators state the same.” (2018 Language Industry Survey, 2018, p. 16)

On the other hand, the Neural Machine Translation did not show any impact on companies or freelance translators to invest in the Neural Machine Translation and develop their own neural engine. Regarding 2018 Language Industry Survey, this is expected in 2019’s survey.

As we can see in the 2018 Language Industry Survey, the translation industry is characterized with positive trends in 2018. Constant developments in the field of the Machine Translation increased the expertise of this industry. “Interesting to note is that quite a few respondents, in particular individual professionals, expect that the lack of quality of the Machine Translation can lead to an increased appreciation for the quality of human translation.” (2018 Language Industry Survey, 2018, p. 29)

On the other hand, the Machine Translation itself remains one of the biggest negative trends connected with the price factor. “The traditional fear that Machine Translation opens the door to lower quality and more competition by lower qualified translators and translation companies remains strong.” (2018 Language Industry Survey, 2018, p. 29)

As we stated before, 2018 is the year of Machine Translation. “On the other hand, it is too soon to conclude that MT is now part of the translation reality, with only some 20% of the language service companies and independent language professionals reporting daily usage. Neural MT has clearly not yet brought the big change that the market is expecting.” (2018 Language Industry Survey, 2018, p. 31)

6. What is Post-editing and Pre-editing of the Machine Translation?

In this chapter, we will elaborate on two terms connected to the field of Machine Translation. One term is the Post-editing and the second is the Pre-editing. We will emphasize their most significant contributions to the field of the Machine Translation and why they are so important these days.

6.1. Post-editing

Lately, the use of Machine Translation software is preferred very often, but the Machine Translation outputs are not yet able to cope with the human translation. In order to achieve a high-quality translation, it is necessary to intervene by the translator or the Post-editor. Translators' Post-editing becomes an increasingly frequent phenomenon for modern translation companies. While the high-quality Machine Translation improves productivity, Post-editing can be very frustrating and often more demanding than the translation process itself. For this reason, an estimation of whether the quality of a translation is sufficient for the Post-editing process is an important area of research.

By Post-editing, we mean another method of evaluating the quality of a translation. Post-editing can be defined as a process in which a Post-editor compares the source text with the target text and modifies it so it meets the purpose for which it is intended. The Post-editing may also be done in monolingual terms, which means that the Post-editor does not have the source text. (Cadwell - Huová, 2016)

It is important to point out that the term Post-editing is used in various fields in the context of processes such as Machine Translation, spell-checkers or even translation memories. The odds are that the profession of a translator and interpreter is not replaceable, because there is no machine that would be able to produce translations in the same quality. In today's fast world, we are all pushed to work as fast as we could. In comparison to the translator is the Post-editing of the Machine Translation faster. It comes to this, that the Post-editing saves not only time but also finances. Simply put, task of the Post-editing is to revise the translated text, which has already been translated by the Machine Translation from the

source language to the target language. However, the Machine Translation is not always correct grammatically or stylistically. (O'Brien, 2014)

We think the importance of Post-editing research is necessary and is underlined by rapid changes in this area. According to Munková, studies on the impact of training and practice in Post-editing will be very essential. All studies that have been published so far have been largely influenced by the fact that the respondents were mostly students who do not have enough experience yet.

An important question remains whether future Post-editors need to be prepared during their studies. Demand for Post-editors is growing as demand for the Machine Translation increases as well. According to Munková, Post-editing should be implemented into the field of study of future translatologists. The ability of high-quality Post-editing will give such translators a better chance in practice.

Regarding Munková, Post-editing should be included in the general education of translators, because this work requires versatility in the field of translation. The Post-editing process itself represents a new situation for the translator that requires specific competencies. As this situation is new, it requires new approaches to education that will meet client requirements in the future. (Munková, 2017)

6.1.1. Types of Post-editing

Post-editing is divided into two types:

- Light Post-editing
- Full Post-editing (O'Brien, 2014)

6.1.1.1. Light Post-editing

Light Post-editing involves minor Post-editing intervention that is required due to the correct meaning of the text. This is done especially when the client needs the translation urgently and in a short time. (O'Brien, 2014)

The Machine Translation outputs are sometimes used without or with little intervention that was applied during the Post-editing process. These texts serve to represent the information and the essence of the text and the target texts have only limited use.

The purpose of Post-editing is to make a text comprehensible and to transfer a background information. The following types of modifications are used during light Post-editing:

- Replacing unknown words
- Removing surplus generated alternatives
- Repairs of meanings in which the machine did not observe the essence of an information
- Correction of the most important words and grammatical errors
- Partial or full rewriting of some sentences

Light Post-editing has several benefits. Most notably, the translation process is simpler and less time consuming than the translation process itself. It is used most often in such communication where the quality of the produced text is not a primary requirement. Most frequently, it is e-mail, simple documents and other types of texts. This type of Post-editing is most often applied, when it is not possible for a human translator to translate a specific text in time.(https://www.rws.com/insights/rws-moravia-blog/bid-353532-Light-and-Full-MT-Post-editing-Explained/, 2019-01-14)

6.1.1.2. Full Post-editing

Full Post-editing means major intervention and editing by the Post-editor in order to achieve a high quality translated text comparable to a human translation. The result is a text that is understandable and appropriately stylistically modified and could be presented for a wider audience. (O'Brien, 2014)

It includes all the steps of light Post-editing and other text modifications such as syntactic or stylistic adjustments. This type of a translation is most often applied when the target text is for public distribution. The produced text must meet the human translation quality standards.

The best quality output in full Post-editing is a text where the recipient does not know whether it is a translated or an original text. It is assumed that the direct translation of a source text is for a translator much easier than the Post-editing. Thanks to the advances that

are being made in this field, it can be changed soon. Nowadays, when the Machine Translation systems use enormous data from specific corpora and databases, clients require form translators to post-edit texts rather than translate. They trust that they will receive a same quality for a lower price. (Cadwell - Huová, 2016)

6.2. Pre-editing

"Pre-editing consists in processing the texts before the Machine Translation. It typically involves correcting mistakes in the source text […], removing ambiguities and simplifying structures. For statistical MT, it may also involve adapting the text in such a way that the input text is a closer match to the texts the engine has been trained with, which can help the Machine Translation engine perform better."
(https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta/, 2019-01-15)

In addition to well-known Post-editing, there is less known Pre-editing. As the name suggests, Pre-editing is the text preparation process before it is translated by a machine in order to increase the output text quality and to reduce the amount of work performed in the Post-editing process. This procedure is most frequently used when the document is translated into more than three languages.

Pre-editing involves many different tasks. One of them is to change the source text so that it uses only short sentences, fixed syntactic structures, exact terms, etc. Other tasks may include spell checking, punctuation, picking the elements that should not be translated and format checking. It is important to note that if the text is to be translated into more languages, it is better to pay attention to the Pre-editing. (http://translation-blog.trustedtranslations.com/do-you-know-the-difference-between-Pre-editing-and-Post-editing-2011-11-23.html, 2019-01-15)

There are still a few more rules to consider when Pre-editing. The important thing is to retain the style of writing and make the work easier not only for the Machine Translation tool but for the Post-editor.

Example for Post-editing:

This text can be found on www.euba.sk. It is a part of the internal directive no.8/2017 on final and habilitation theses.

Source sentence without Post-editing:

Okrem vloženia elektronickej verzie do systému evidencie záverečných prác odovzdá študent, resp. uchádzač habilitačné konanie vytlačenú formu záverečnej alebo habilitačnej práce.

Translation:

In addition to inserting an electronic version into the system of registration of final works, the student, or. applicant for habilitation procedure printed form of final or habilitation thesis.

Source sentence with Post-editing:

Študent, resp. uchádzač o habilitačné konanie vloží elektronickú verziu do systému evidencie záverečných prác. Okrem toho sa odovzdáva aj vytlačená forma záverečnej alebo habilitačnej práce.

Translation:

Student, respectively applicant for habilitation procedure will put the electronic version into the system of registration of final works. In addition, a printed form of final or habilitation work is also submitted.

In the translation above, we have shown an example of Pre-Edited Machine Translation. Without Pre-Editing, major errors occur in the translation and the understanding of such translation is not possible. With appropriate Pre-editing, we can avoid major shifts in meanings and improve the output translation. By simplifying the source text, we have made it easier for the Machine Translation to evaluate and reduce the error rate.

7. Objectives and methods of the thesis

In the following chapter, we will describe the main objective of this thesis as well as following hypotheses. We will focus on the fact, if these hypotheses will support the main objective of this thesis or not.

Moreover, we will characterise the research methods of this thesis, which were used in order to create a sufficient result with a reliable information value. More precisely, we will try to make our own forecast of the future of the translation industry not only in Slovak market, but globally as well.

7.1. Objectives of the thesis

We live in a society where we can meet people from all over the world and with these people, we probably could not communicate in a way in which we would like to. Unfortunately, we will never know all the languages that people communicate currently. That is the time when we look for someone who enables us to communicate with such a person or something what serves and helps us to understand what is a particular person saying. We more and more often decide for the Machine Translator to support us.

The general objective of this thesis is to analyse the frequency of the usage of the Machine Translation tools by the freelance translators and the translation agencies in Slovakia. We will try to find out whether is the growing trend of the usage of the Machine Translation among the wide audience the same also among professional translators and translation agencies in Slovakia. After the analyses, it will be easier to see if the Machine Translation has its own space in the context of translating not only when used by wide audience but also by professionals.

We have formulated several hypotheses on which we will elaborate later. Hypotheses were created to support the main objective of the thesis. Those are as follows:

- **Hypothesis 1**: An average translator has been involved in translation services for over 10 years.
- **Hypothesis 2**: An average translator works on 10 or less translation projects a month.
- **Hypothesis 3**: The most frequently translated texts by the translators are legal texts.
- **Hypothesis 4**: English language and Slovak language are the most frequently source and target languages in Slovakia.

- **Hypothesis 5**: For the translators it is more efficient to translate a text with the Machine Translation tool and then Post-Edit it.
- **Hypothesis 6**: Google Translate is the most frequently used the Machine Translation tool by the translators.
- **Hypothesis 7**: The Neural Machine Translation is qualitatively better for translators than the other types of the Machine Translation.

7.2. Research methods of the thesis

The questionnaire is one of the most frequent used research methods. It is used in social sciences to collect and quickly identify facts, opinions, attitudes, preferences, values, motives, needs, interests, and so on.

We used quantitative research to find out the required information. The source of information in the research section was, as we mentioned above, the online questionnaire. Completing the questionnaire was anonymous and voluntary. The questionnaire was structured and contained 18 questions.

Questions 1, 2, 9, 10, 13, 16, 17, 18 are closed. Respondents had the opportunity to choose one of several options that best reflected their opinion. Some of that questions were evaluating questions from 1 to 5 (1 is the best and 5 is the worst).

Questions 3, 4, 14 are open and allowed respondents to express their own opinion on the questioned issue.

In questions 5, 6, 7, 8, 11, 12, 15, respondents had the option of choosing more than one option to express their opinion more specifically.

Before we started researching, we accomplished a literature search. Having studied the current problematic of the topic, we had enough information to compile a questionnaire that would help us to summarise the most important information in the translation industry. We distributed the survey via a web-based questionnaire that was posted on social networks or sent in a private communication. Promotion of the questionnaire was done by messages to chat groups, page administrators, information in e-mails and orally as well.

The respondents had the opportunity to complete the questionnaire in the second half of 2018.

46 respondents completed the questionnaire. Of the total number of respondents, 25 (56,5%) were women and 20 (43,5%) were men. In this case, it is not possible to calculate the return of our questionnaire, as the extent and characteristics of the respondents were not known.

7.3. Research questions

At the beginning of the questionnaire, we chose the following questions:

- **Question 1**: Gender.

This question is focused on the gender of respondents who participated in the survey.

- **Question 2**: How long have you been professionally engaged in translation?

The question was to find out how long respondents have been in the translation industry and whether the respondents are already experienced translators or just newbies.

- **Question 3**: How many projects do you translate a month?

This question centred on the number of translations that translators could translate within one month (20 business days).

- **Question 4**: How many projects do you translate a year?

As with the previous question, we tried to find out the workload of the respondents.

- **Question 5**: What are the key areas of your translation work?

This question is addressed to the content of the texts that translators most often work with.

- **Question 6**: What kind of texts do you devote to your work?

This question is related to the type of texts translated by the translators most frequently.

- **Question 7**: Which source languages do you mostly translate from?

This question is focused on which languages the translators most often meets.

- **Question 8**: Which target languages do you mostly translate into?

This question is addressed to the languages into which translators translate most frequently.

- **Question 9**: Do you use the Machine Translation for your translations?

This question is aimed at the fact, if the translators use the Machine Translation tools in their translations.

- **Question 10**: How often do you use the Machine Translation tools while translating?

Question 10 is related to the frequency of the use of the Machine Translation tools by the translators.

- **Question 11**: Which of the freely available Machine Translation tools do you use most often?

This question is aimed at the type of the freely Machine Translation tools used by translators.

- **Question 12**: Which of the commercial Machine Translation tools do you use most frequently?

This question is focused on the type of the commercial Machine Translation tools used by the translators.

- **Question 13**: If you use any of the available Machine Translation tools, how do you evaluate the quality of a particular Machine Translation tool.

This question is dealt with the fact, how the translators evaluate a particular Machine Translation tool, if they use any, from their point of view.

- **Question 14**: If you do not use any of the available Machine Translation tools, what are the reasons?

In this question were translators asked to write several reasons why they do not use any of the available Machine Translation tools.

- **Question 15**: Which type of a translation do you think is the most effective for the translator in terms of time?

This question is aimed at the most effective way for the translators to translate.

- **Question 16**: Have you ever heard of the concept of the Neural Translation in your work?

This question is related to the fact, if the translators have ever heard of the Neural Translation.

- **Question 17**: Do you use any software based on Neural Translation in your work?

This question is focused on the engine of the software used by the translators.

- **Question 18**: If you use any software based on the Neural Translation, how would you evaluate it?

This question is focused on the evaluation by those translators, who use software based on the Neural Machine Translation method.

8. Results of the thesis, discussion and future recommendations

In the next chapter, we will take a closer look at the results of our research, which consisted of a questionnaire distributed last year and filled in by the freelance translators and the translation agencies in Slovakia. In the previous chapter of our diploma thesis, we explained the research methods and specified the questions in the distributed questionnaire. The questionnaire was completed by both experienced and less experienced translators of different age groups.

In order to better achieve the stated objective of our diploma thesis, we have compiled seven hypotheses that will help us to better understand the behaviour of the translators and the translation agencies in their translation services. We will also try to characterise the average Slovak translator in terms of experience, type of translated text, and frequency of using the Machine Translation tools. We will analyse each hypothesis individually based on the answers of the respondents.

We will try to compare the results of the hypotheses with the results of the 2018 Language Industry Survey, which is discussed in more detail in Chapter 5, and we will try to identify similarities or differences between the Slovak translation industry and the worldwide one.

8.1. Results of the thesis

Hypothesis 1: An average translator has been involved in translation services for over 10 years.

The first hypothesis focuses on whether translators in Slovakia have been translating for more than 10 years in average. With this hypothesis, we wanted to find out if they are translators that are more experience or newbie translators. The graph shows the years during which translators have been actively engaged in their profession. The results are as follows:

30.4% (14) of respondents spend 5 years or less on translation industry. We can assume that they are less experienced and novice translators or students of translation disciplines.

23.9% (11) translators spend from 6 to 10 years on translation services. We assume that they are translators who have just completed their university studies, or are more

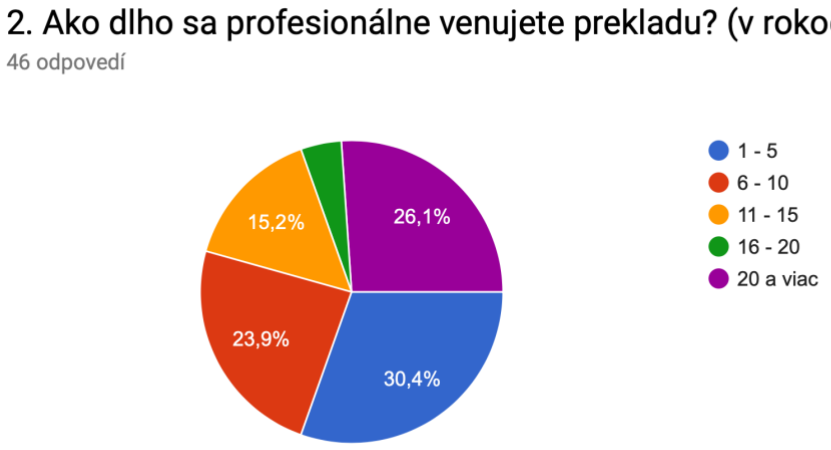
experienced translators who have already begun to do translation activities during their studies, or have decided to start translating without studying at university at all.

15.2% (7) of respondents are in the translation business from 11 to 15 years. We think they are translators who already have a professional background and reputation as well as connections in this translation industry. 4.3% (2) translators are dedicated to their translation activities for 16 to 20 years.

26.1% (12) of people stated they have been translating for more than 20 years. In this case, they are already experienced translators with many years of experience, who have already translated many various texts.

The graph shows that 54.3% (25) of respondents spend less than 10 years in the field of translation. 45.7% (21) translators are part of the translation industry for over 10 years.

We have refuted our first hypothesis. The reason for this may be a strong proportion of translators who have been involved in the translation activity for 1 to 5 years, so starting translators or students of translation.



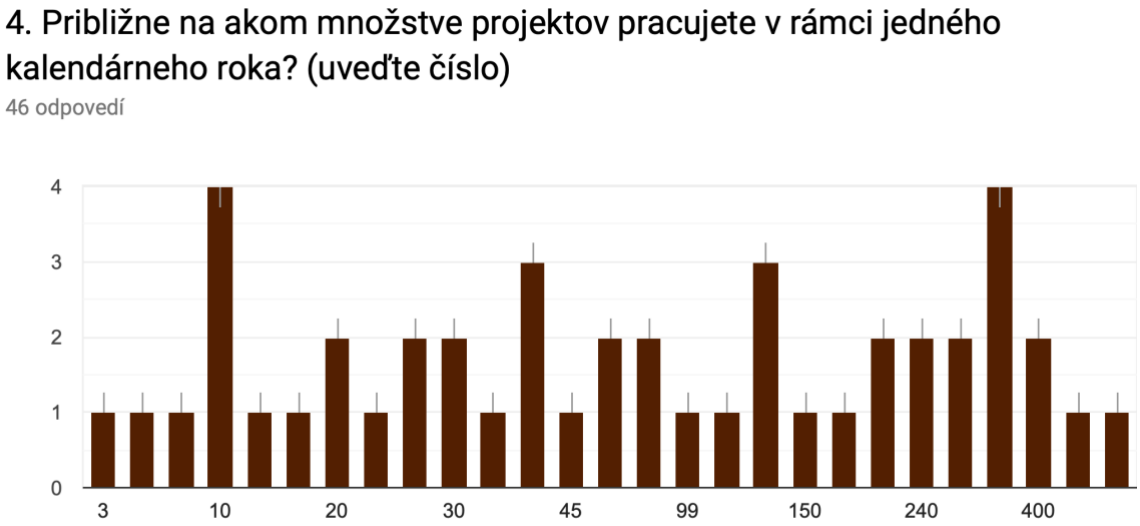
Graph 1: How long have you been professionally engaged in translation?

Hypothesis 2: An average translator works on 10 or less translation projects a month.

The second hypothesis focuses on the number of translations performed by the translator during one month. We focused on comparing two questions from our questionnaire, more specifically question numbers 3 and 4.

Question 3 identified the number of projects within one calendar month and question 4 quantified the number of translation projects within one year. We will compare these two values and try to establish a direct link between the number of translations during one month and the number of translations during one year. One question controls the other and vice versa.

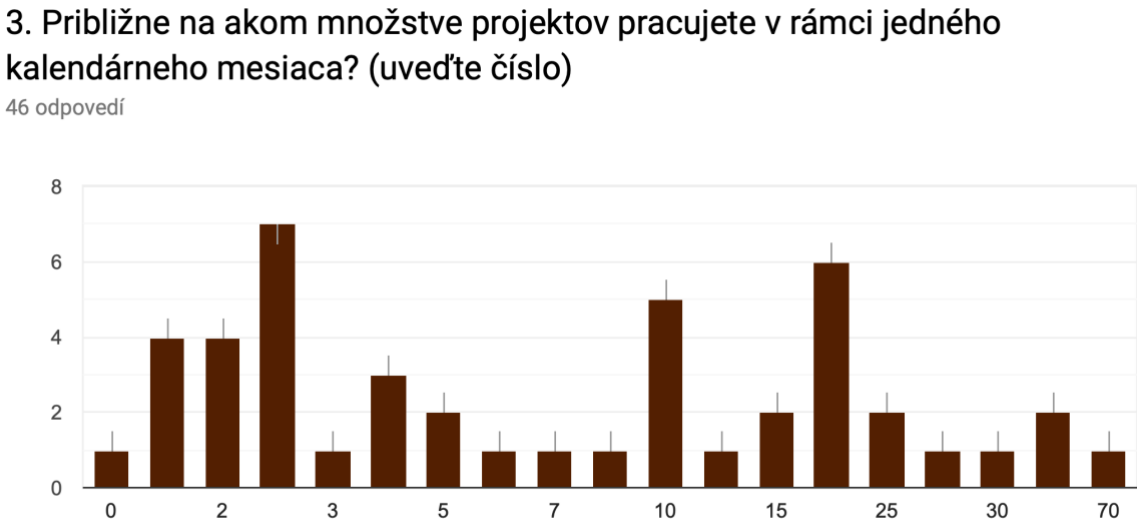
The graph below and question 4 showed that the average annual number of translation projects is 105.3. This means that the average translator works on 105 translations in one year, indicating that the average translator can handle 8.7 translations in one month. Question 3 clarified that the translator can translate an average of 8.4 translations per month.



Graph 2: How many projects do you translate a year?

Both questions show a similar number of translation projects in one month. Thanks to two similarly focused questions, we can easily verify the accuracy of the data and evaluate their truthfulness.

54.3% (25) of respondents stated they work on less than 10 translation projects a month. 45.7% (21) of the interviewed translators said they translate more than 10 translations or translation projects per month. Our second hypothesis was confirmed with the result that the average translator works on 8 to 9 projects in one month. More detail can be seen in the graph below.



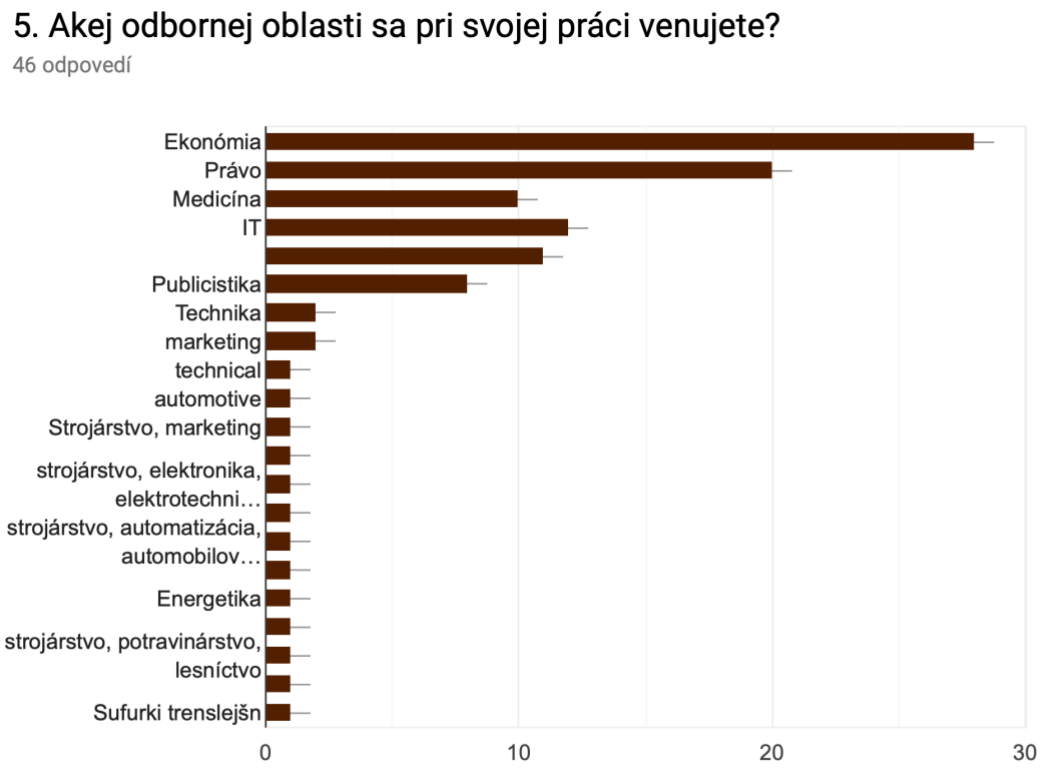
Graph 3: How many projects do you translate a month?

Hypothesis 3: The most frequently translated texts by the translators are legal texts.

In the third hypothesis, we focus on what type of text translators most frequently translate in Slovakia. To quantitatively achieve these results, combination of two questions is necessary. More specifically, we are combining question 5 and question 6.

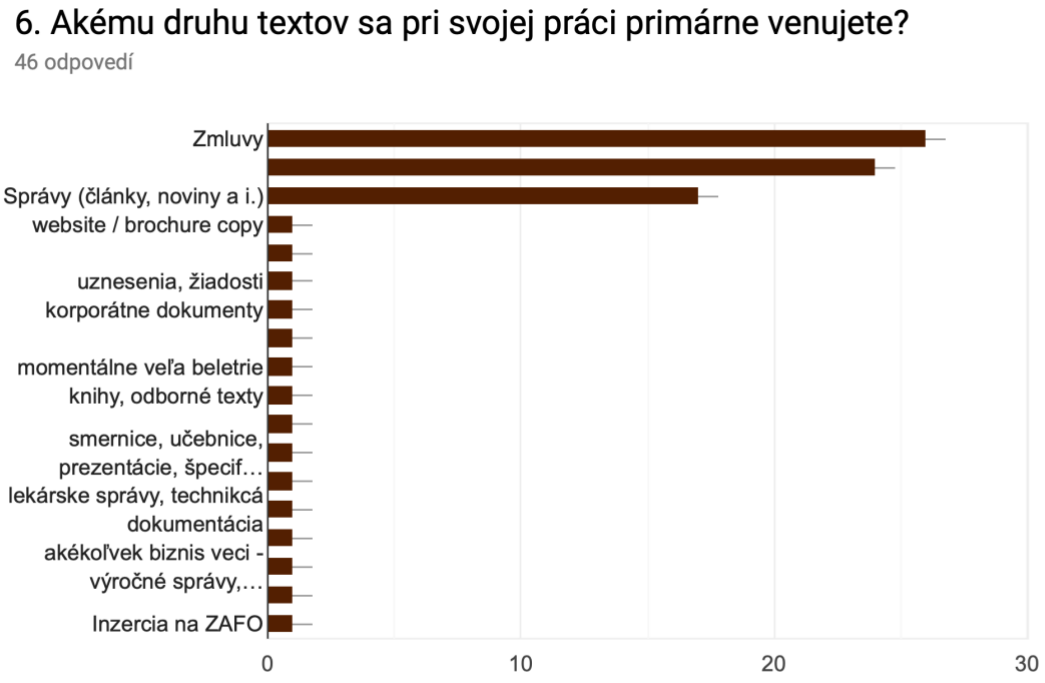
Question 5 specified a particular type of texts translated by professional translators in their projects. Question 6 observed the sort and characteristics of texts most frequently translated by our respondents. There was a choice of multiple answers on both questions, so the final numbers are proportionally different.

The graph below shows that 28 respondents focus in their work on economic texts. 20 translators work in the field of law in their business. 10 respondents are engaged in translating medical texts, 12 respondents translate IT texts, and 11 translate fiction texts. 8 respondents deal with translation of journalistic texts. Other categories such as engineering, forestry, electronics, marketing are represented only in minimal or negligible quantities.



Graph 4: What are the key areas of your translation work?

The graph below illustrates the type and characteristics of texts most frequently translated by professional translators. 26 translators chose contracts as the most frequently translated type of texts. 24 respondents said they most often translate guides and manuals. 17 respondents chose reports (newspapers, articles) as the most frequently translated type of texts.



Graph 5: What kind of texts do you devote to your work?

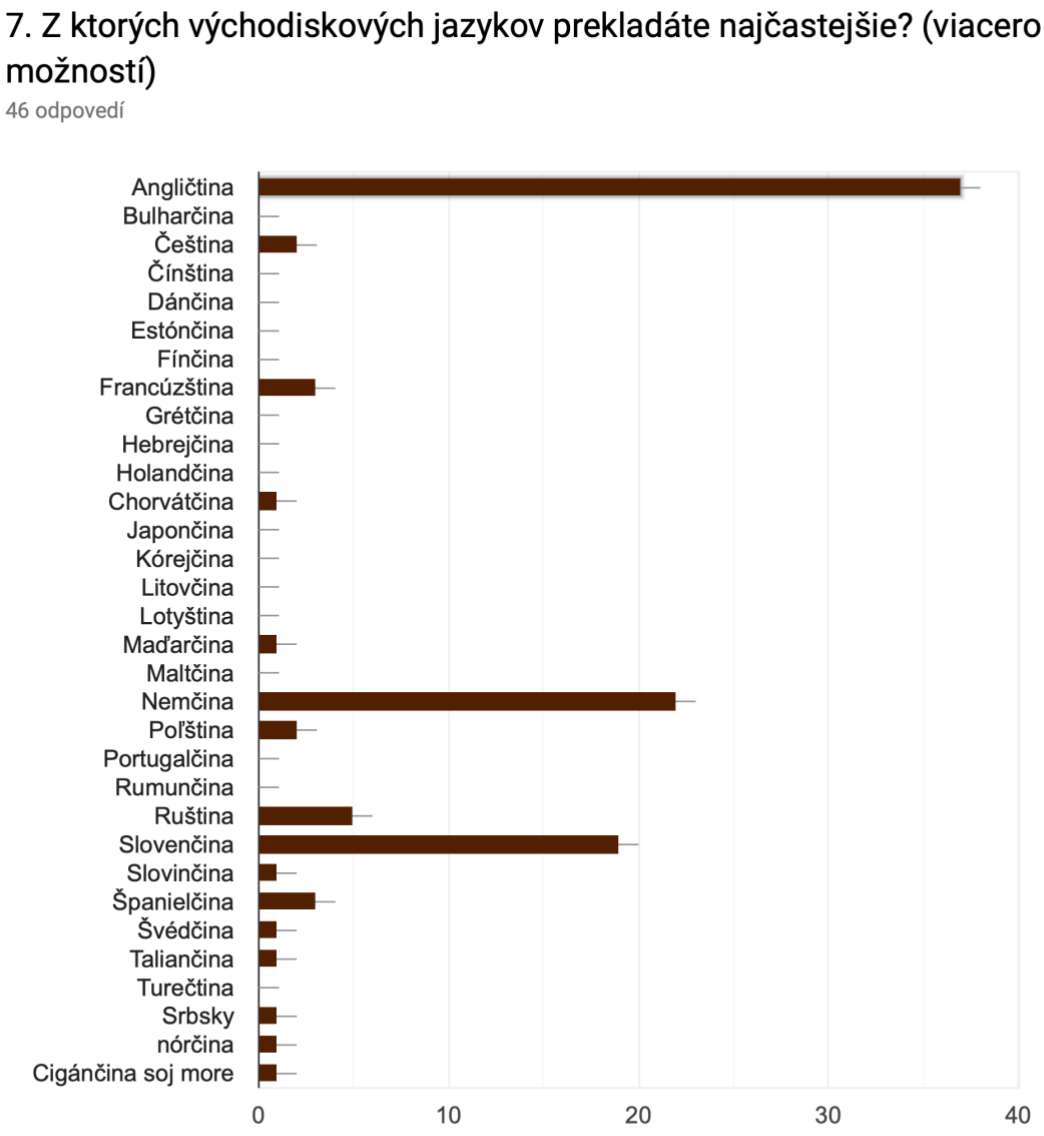
To better understand and process the results of the questionnaire, we need to look at the individual responses of our respondents. With 20 replies, translations of legal texts were placed second. On the other hand, translations of economic texts with 28 responses ranked first. Thus, economic texts are the most frequently translated texts in the Slovak translation industry. We have refuted our hypothesis. In comparison to the world market, the Slovak market is slightly different.

Hypothesis 4: English and Slovak language are the most frequently source and target languages in Slovakia.

Our fourth hypothesis focuses on which languages the translators most frequently use in their translation services. We assume that the most commonly used languages are English and Slovak. In the questionnaire, we have chosen a combination of two questions for this hypothesis.

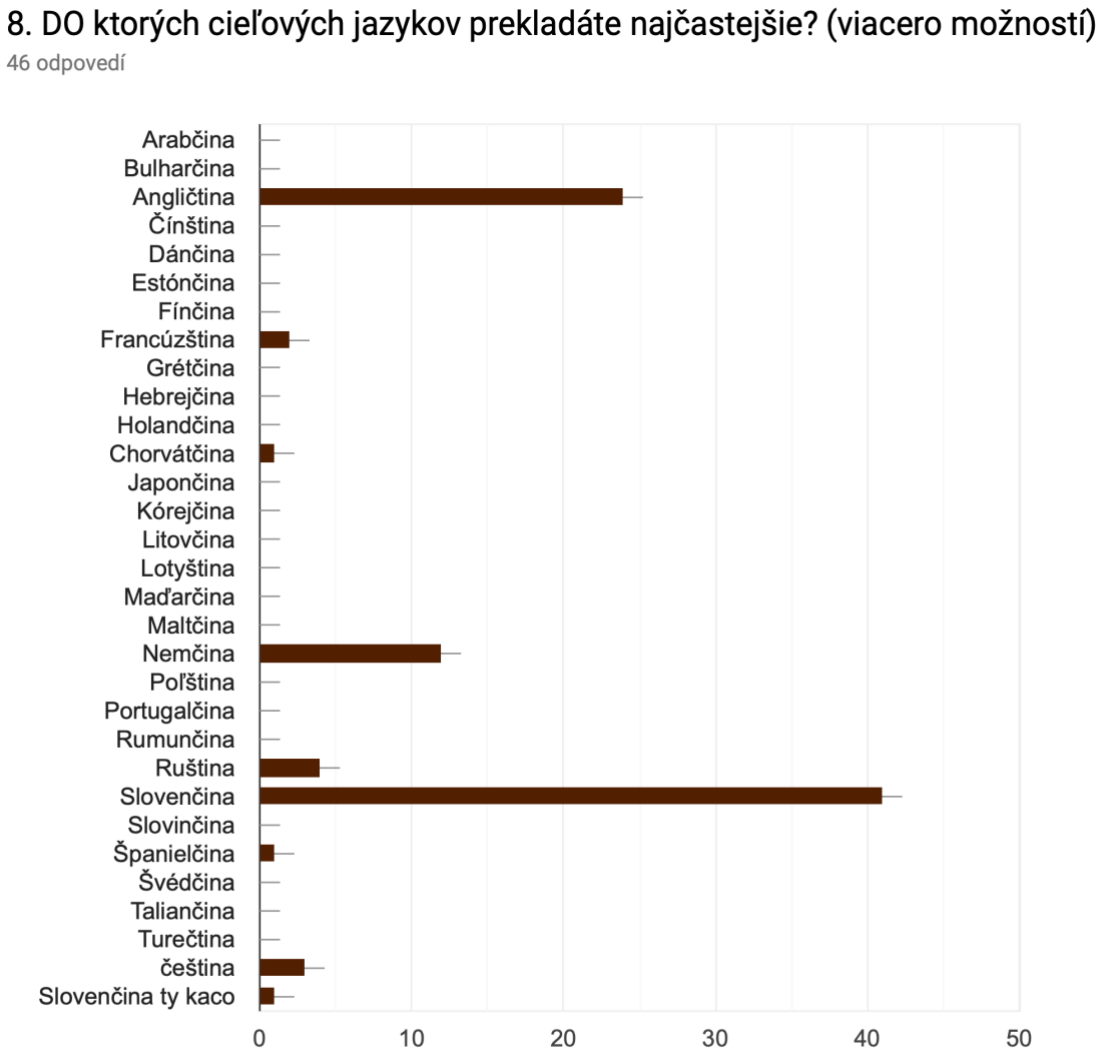
In question 7, the respondents expressed what are the most common languages used by the professionals in translations. Question 8 is focused on which target languages are most frequently used by translators.

In the attached graph of question 7, we can see that 37 people chose English as the most widely used source language. German is second, selected by 22 translators. 19 translators have chosen Slovak as the most commonly used source language. Languages such as Russian, Spanish, French and Czech have been chosen only in minimal quantities and are therefore not relevant.



Graph 6: Which source languages do you mostly translate from?

The graph below shows to which target languages the translators in Slovakia most frequently translate. Up to 41 translators have chosen Slovak and thanks to this we can see which target language is the most often requested from the translators. 24 translators chose English and German chose 16 experts. Other languages such as Russian, Czech and French are irrelevant to the result of our objective.



Graph 7: Which target languages do you mostly translate into?

The result of our closer research is the fact that 37 respondents chose the English-Slovak language combination in our questionnaire. 24 translators selected English-German language combination. From the mentioned above results that the most commonly used

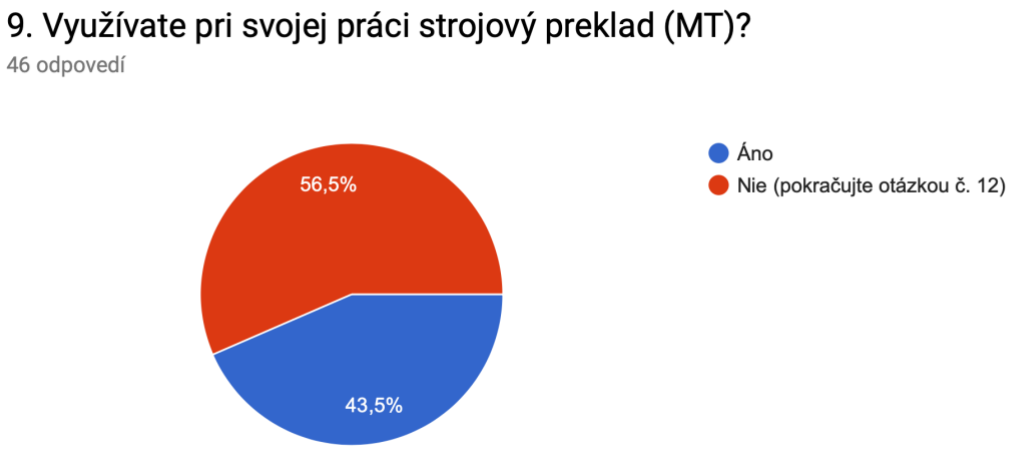
language combination for the source and target languages is English and Slovak, respectively Slovak and English. Our fourth hypothesis is therefore confirmed thanks to these results and these two languages are the dominant language combination on our market.

Hypothesis 5: For the translator it is more efficient to translate a text with the Machine Translation tool and then Post-edit it.

Tackling the fifth hypothesis, five questions have to be discussed. More specifically, these are questions 9, 10, 13, 14, 15. This hypothesis focuses on whether it is more efficient for the translators to use Machine Translation tools and then to Post-edit such a translation, or to rely solely on their capabilities and skills throughout the whole process of translation.

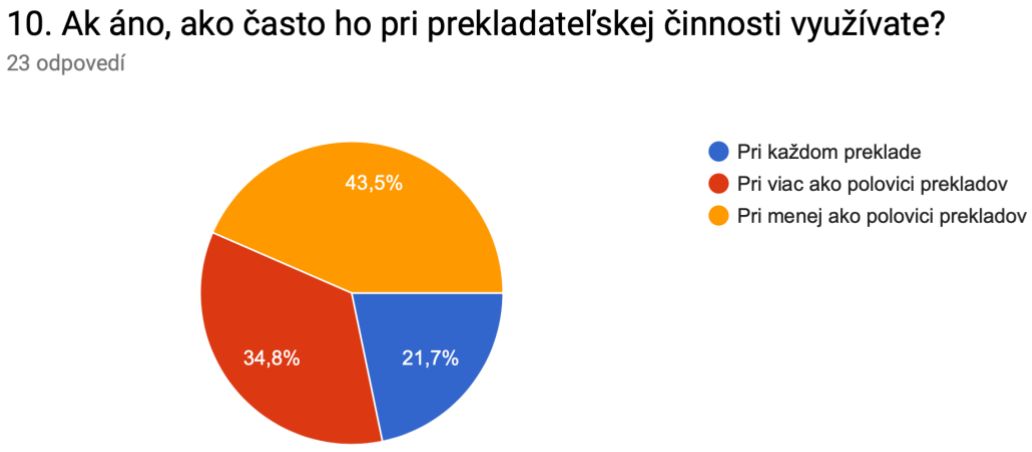
The first question we watched is whether the translators use the Machine Translation tools in their work. The worldwide trend in the frequency of use of the Machine Translation by the translators is increasing. According to the 2018 Language Industry Report, more than a half of the translators are already using the Machine Translation. Research has shown that 2018 was the year when the Machine Translation is used for more than a half of the translations (2018 Language Industry Report).

The graph below shows the ratio of the translators who use the Machine Translation for their services. On the other hand, there are translators who rely solely on their language competences, abilities and skills. In our questionnaire, 43.5% (20) of respondents said they are actively using the Machine Translation tools in their work. Surprisingly, 56.5% (26) of the translators said they are not actively using the Machine Translation services in their work. In this aspect, the Slovak market has different trends in the development than the world market.



Graph 7: Do you use the Machine Translation for your translations?

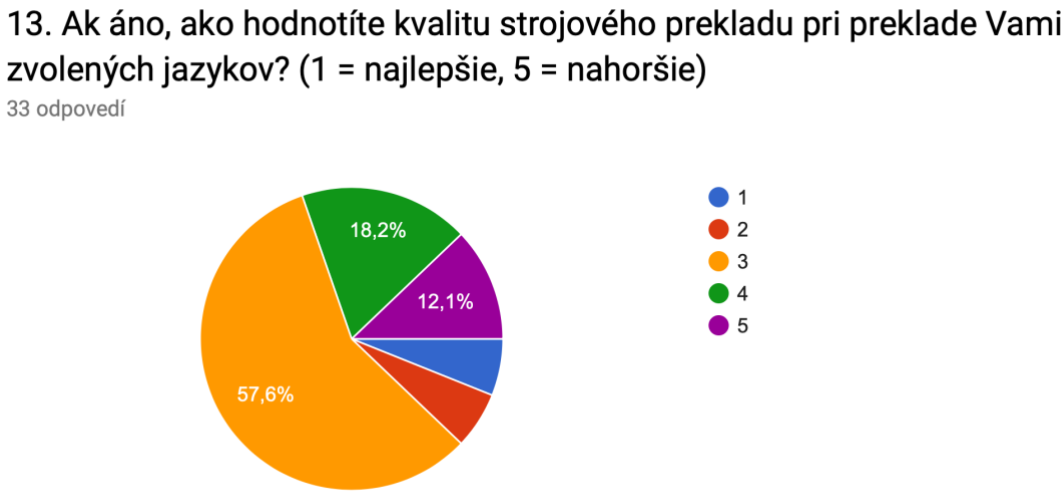
In question 10, we investigated how often the translators use the Machine Translation services in their projects. Research has shown that 21.7% of respondents use the Machine Translation for each translation. 43.5% of translators say they use the Machine Translation tools in less than a half of their translations. 34.8% of respondents chose that they use the Machine Translation tools for more than a half of the translations. In the graph below, we can see the distribution of each category.



Graph 8: How often do you use the Machine Translation tools while translating?

In question 13, respondents expressed their satisfaction or dissatisfaction with the quality of the Machine Translation. They were able to evaluate the Machine Translation on

a scale of 1 to 5, where 1 was rating for the best translations and 5 for the worst. 30.3% of respondents said they were dissatisfied with the quality of the Machine Translation. The Machine Translation was rated with 5 and 4. On the other hand, 12.1% of translators said they were satisfied with the quality of the Machine Translation. The Machine Translation was graded with 1 and 2. The largest group of respondents (57.6%) rated the Machine Translation with number 3, which means that they did not express either direct satisfaction or explicit dissatisfaction. We can assume that they use the Machine Translation as a support tool in their translation projects.



Graph 9: If you use any of the available Machine Translation tools, how do you evaluate the quality of a particular Machine Translation tool.

Our next question was about why translators refuse to use the Machine Translation tools. In this question, translators had the opportunity to list concrete reasons why they do not use the Machine Translation. Moreover, we chose the most important ones and the most frequent ones.

The following are the most frequent reasons:

- inaccuracy
- time-consumption
- low efficiency

According to the translators, these three reasons are the most frequent obstacle to not use the Machine Translation tools.

Equally important are the reasons we have decided to highlight and emphasise:

- inapplicability in artistic texts

- confidentiality
- shifts in context

Among the most interesting reasons we decided to include that, according to some respondents, the use of the Machine Translation tool is not behaviour of a professional and educated language expert. Another compelling reason not to use the Machine Translation is the change in a meaning of a particular part of the texts, which can cause great problems when submitting the translation to the client.

Question 15 was a key question for answering our fifth hypothesis. We asked respondents what kind of translation is the most effective in terms of time. They were able to choose from options such as the Machine Translation, the Computer-Aided Translation and the Post-Edited Machine Translation. Answers to this question had the highest priority when evaluating the fifth hypothesis. The results are discussed below. Translators have the option to select multiple responses. This fact helped us to determine how many translators will choose a combination of translations followed by Post-editing.

15.2% (7) of respondents said that the Machine Translation is the most appropriate way of translation in terms of time. 76.1% (35) respondents chose the Computer-Aided Translation as the most effective way of translation. 10.9% of respondents (5) chose Post-Edited Machine Translation as the most appropriate way to translate. If we take a closer look at the results, we found that the vast majority of translators rely on Computer-Aided Translation. We have refuted our hypothesis and this particular hypothesis is in contrast to our assumption as well as the hypothesis, where we discussed how translators evaluate the quality of the Machine Translation. Below is a graph with results in more detail.

Graph 10: Which type of a translation do you think is the most effective for the translator in terms of time?

Hypothesis 6: Google Translate is the most frequently used Machine Translation tool by the translators.

Our sixth hypothesis focuses on which tool from commercial or freely available translators is most frequently used by the translators. In the following lines, we will analyse two questions from our questionnaire and find out which translation tool is the most popular.

Question 12 focused on various types of commercial translation tools. 33.3% of respondents said that they most often use Microsoft Translation Hub. Other responses were distributed among different translation tools and had no impact on the overall result. Among the chosen translation tools were Omniscien, SDL BeGlobal, Systran, Babylon and SDL Trados. Several respondents have added that they use a system provided by the client. Below is a graph with specific results.

Graph 11: Which of the commercial Machine Translation tools do you use most frequently?

Question 11 focused on freely available translation tools. 90% of respondents clearly chose Google Translate as the most commonly used Machine Translation tool. 10% of respondents said Microsoft Bing Translation was the most frequently used translation tool. Below we can see the graph with answers in more detail.

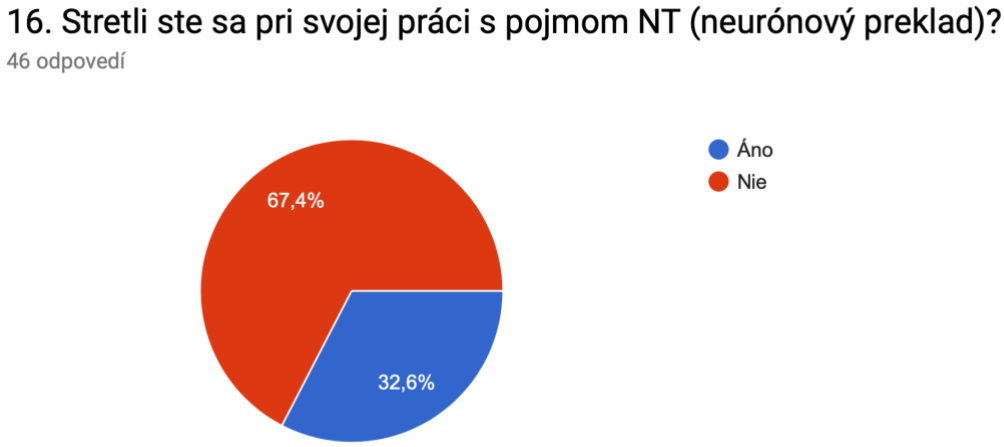
Graph 12: Which of the freely available Machine Translation tools do you use most often?

It follows from the above answers that Google Translate is the most frequently chosen type of translation tool among respondents. Thanks to these answers, we have confirmed our hypothesis and we see a growing influence of Google Translate among translators.

Hypothesis 7: The Neural Machine Translation is qualitatively better for translators than the other types of the Machine Translations.

Our seventh hypothesis focused on whether the translators who understand the use of neural translation are qualitatively better than other types of Machine Translation. We will address three questions from our questionnaire, namely questions 16, 17 and 18.

In Question 16, we were wondering whether the translators had already actively encountered the term neural translation. 32.6% (15) of the respondents said they had encountered the concept of neural translation in their work. On the other hand, 67.4% (31) of respondents said that they did not meet actively in this way of translation. Compared to the situation in a world where neural translation is becoming increasingly important, this term is still relatively unknown in Slovakia. However, it is more than likely that we will hear more and more about this type of Machine Translation. We see the results in the attached graph.



Graph 13: Have you ever heard of the concept of the Neural Translation in your work?

Question 17 focused on whether translators are actively using software that supports neural translation. 88.6% (39) of the respondents said they did not use neural translation support software in their activity. On the other hand, 11.4% (7) of respondents said they were actively using software supporting neural translation. Looking at the fact that in the question of 11 to 90% of respondents said they were using Google Translate, we conclude that users really did not meet the concept of neural translation. Google Translate has been promoting neural translation for years and it is very interesting to see such a contrast. Below we can see the respondents' individual responses.

Graph 14: Do you use any software based on Neural Translation in your work?

Question 18 focused on how users of tools supporting the Neural Translation evaluate this Machine Translation method. 12.5% of respondents expressed dissatisfaction with this way of translation, which is a great surprise for us. 37.5% of the respondents expressed a neutral attitude towards this translation method. These respondents expressed neither satisfaction nor dissatisfaction. On the other hand, 50% of respondents expressed satisfaction with the Neural Machine Translation and the quality of the resulting text. Thanks to these responses, we can conclude that compared to other types of the Machine Translation, the Neural Machine Translation is rated better. We have confirmed our hypothesis and we see that the quality of the Machine Translation has moved to a higher level thanks to the application of neural networks. The following graph shows the responses of our respondents.

Graph 15: If you use any software based on the Neural Translation, how would you evaluate it?

8.2. Discussion

The general objective of our thesis was to find out how translators and translation agencies use the Machine Translation in Slovakia. More precisely, the aim of our diploma thesis was to focus on the frequency of using the Machine Translation by translators and translation agencies. Nowadays, a large number of private as well as legal persons opt for the services of a professional intercultural mediator who, through their linguistic skills, allows them to offer their services or goods beyond the borders of their country of origin. As a result of this, the market in the world is globalising and currently, there is no problem shopping outside our home country. Thanks to the translators and the Machine Translation, companies have a better coverage on much more potential customers which could be reached easily and offered with new goods and services.

In our diploma thesis, we tried to find out how the translators and the translation agencies perceive the term Machine Translation and how they use it in their translation activities. Moreover, we tried to quantify the use of the Machine Translation in commercial translation in Slovakia. For the purposes of research and exact analysis, we have compiled a questionnaire with 18 questions that helped us to quantify the current situation of the Slovak translation market. For the overall evaluation of the general objective of our diploma thesis, we decided to compile 7 hypotheses, the results of which we used to analyse the current state

of commercial translation in Slovakia. The questionnaire showed very interesting results, which will be specified below.

Our first hypothesis "An average translator has been involved in translation services for over 10 years." was refuted. Based on this result, we can see that there is a predominance of translators who are newcomers to this industry in Slovakia. On the other hand, a relatively large group of translators said they had been working in translation services for over 20 years. We understand that the translation market in Slovakia is very diverse, but also that there is still a great deal of interest from young people and students in the translation industry in our country. Therefore, translation services are still an attractive profession in Slovakia, despite some claims that the translator will be replaced by a machine over time. According to the 2018 Language Industry Report, the biggest threat to translators is currently reducing translation prices. We agree with this statement. On the other hand, the trend of studying translation in Slovakia is positive and there will be no shortage of qualified translators in the future. (2018 Language Industry Report)

The second hypothesis of our thesis "An average translator works on 10 or less translation projects and a month." was confirmed. The Slovak average translator works in 8 to 9 translation projects a month. In our opinion, this is an absolutely reasonable number regarding the highest possible quality of a translation. An important fact for us is whether the translator actually uses translation tools such as the Machine Translation or the CAT tools. Using the translation tools, the efficiency of the translator could be increased. Only the output quality of the translation remains questionable. This aspect should be an individual preference for each language expert.

The 2018 Language Industry Report states that legal texts are the most frequently translated type of a text. In our third hypothesis, we wanted to confirm or refute this statement in terms of the Slovak market. Our questionnaire revealed that 20 respondents decided to translate legal texts. On the other hand, 28 respondents decided that they most often translate economic texts. Thanks to this question, we have refuted our third hypothesis and, compared to the 2018 Language Industry Report, we see a difference in the behaviour of the translators in Slovakia and the world. According to this survey, the greatest number of translators are clearly dedicated to translating legal texts. (2018 Language Industry Report)

In the fourth hypothesis "English and Slovak are the most frequently source and target languages in Slovakia." we confirmed that the most frequently translated source and target languages are English and Slovak, respectively Slovak and English. English is

currently the language number one in the world. In our opinion, nothing will change this fact, and English will be the language of globalization. It is understandable because the market in the west is more interesting for companies as well as for the translators than the market in the east.

Fifth Hypothesis "For the translator it is more efficient to translate a text with the Machine Translation tool and then Post-edit it." focused on how the translation process of a translator or translation agency looks. It is common practice in the world that clients translate text through the Machine Translation or the CAT tools. Translators are required only to proofread or Post-edit such a translation. In Slovakia, a large majority of translators rely on the CAT tools. In our opinion, the global trend of Post-Edited Machine Translation will be, soon or later, reflected in our translation market. We agree with the modern approach that it is more efficient for the translator to Post-edit the translated text rather than translate the entire text on their own and based on their skills and abilities. Many translators disagree with this opinion, but the trend of this development is undeniable.

Our Sixth Hypothesis "Google Translate is the most frequently Machine Translation tool by the translators." focused on whether Google Translate is the most used translation tool in commercial translations in Slovakia. According to the 2018 Language Industry Report, Google Translate is a clearly dominating translation service among all freely available translation tools as well as commercial translation tools. (2018 Language Industry Report) The third chapter explains that Google has implemented neural translation into its translation tool, which has undoubtedly improved the quality of this online service. For us, Google Translate has been immensely improved in terms of quality and overall impression of the output translation. The fact that this online translation tool is freely available will exceedingly contribute to its faster development and progress in the upcoming years.

The seventh hypothesis "The Neural Machine Translation is qualitatively better for translators than the other types of the Machine Translations." has explained to us whether the translators are satisfied with the quality of the neural translation itself and whether they evaluate it better than the Machine Translation. We think that the quality of the neural translation is much higher than the quality of the Machine Translation. This fact was confirmed in our hypothesis as well. Neural translation meets a lower translation error rate and therefore this type of Machine Translation is certainly the most interesting for developers and translators in the coming years.

8.3. Future recommendations

The issue of the Machine Translation has become a very important aspect in the translator's work process since its origin. This technology has become a common part of everyone's life. Based on the results of the questionnaire, we have several future recommendations as well as recommendations for practice. These recommendations can be divided into two groups:

- recommendations for experienced translators
- recommendations for beginning translators

It is important to look at the fact how are experienced translators taking the Machine Translation into consideration. Our results show that such experienced translators rely on their long-term skills and abilities. In the current situation, however, even they should not ignore phenomena such as the Machine Translation, the Neural Machine Translation and others. In the upcoming years, these tools will come to the forefront and their position will be irreplaceable. Experienced translators who do not use the Machine Translation tools should certainly give them a chance. It is a fact that many experienced translators do not identify themselves with the Machine Translation phenomenon, but thanks to such tools they can increase the efficiency of their translations, reduce the time consumption and thus satisfy more clients.

On the other hand, we have a relatively large number of newbie translators, who are less experienced, but born into the world full of technology, which they quickly got used to. With this aspect, they are largely familiar with the Machine Translation and various other tools used on a daily basis. For the vast majority of these translators, the global trend of the Machine Translation and other CAT tools is a very interesting topic. With the implementation of new systems and technologies, the Machine Translation tools are developing very quickly. Therefore, we encourage newbie and less experienced translators to be actively interested in the Machine Translation tools and other translations supporting tools.

Conclusion

In recent years, the Machine Translation has become one of the most important aspects of the translation industry worldwide. Its importance is increasing even more thanks to international globalization and the growing dependence of states on one another. In our diploma thesis, we tried to define the frequency of using the Machine Translation in commercial translation in Slovakia. More specifically, we tried to determine, in addition to the frequency of the Machine Translation, how translation entities in Slovakia perceive the concept of the Machine Translation and how they deal with it.

In the first chapter, we focused on the characteristics and the origin of the term Machine Translation. History and origin of the Machine Translation date back to the 1960s. In 1954, the first public Machine Translation test took place. This experiment was made possible thanks to the collaboration between IBM and Georgetown University. This successful attempt meant a new beginning or a renaissance for the translation industry. After 10 years of research, the initial enthusiasm has disappeared and the concept of the Machine Translation has almost completely fallen into dust. Furthermore, in the first chapter, we focused on the categorisation of the Machine Translation itself and we described the beginnings of the Machine Translation in Slovakia.

In the second chapter, we talked about the process and methods of the Machine Translation. We can say that there are two basic approaches in the Machine Translation. These translation methods are:

- Rule-Based Machine Translation (RBMT)
- Corpora-Based Machine Translation (CBMT)

Each of these approaches has its own specific characteristics. More specifically, we also talked about how the different approaches are running.

In the third chapter, we dealt with the specific Machine Translation tools. We have taken a closer look at Google Translate and the Systran translation tool. We have characterized them from our point of view. We have specified their history and how both translation tools work.

In the fourth chapter, we summarised the latest Machine Translation method, namely the Neural Machine Translation. As mentioned in the second chapter, we can classify the Neural Machine Translation as one of the Corpus-Based Machine Translation tools. This Machine Translation approach is currently one of the best. The fact is that even this type of

the Machine Translation has its disadvantages, but on the other hand, its error rate is much lower than with the other Machine Translation approaches. Moreover, we compared the Neural Machine Translation with the Statistical-Based Machine Translation.

In the fifth chapter, we have demonstrated the most frequent errors of the Machine Translation from a several points of view. Specifically, it was in terms of common communication, context impact, polysemy, synonyms, homonyms, idioms and dynamic language. On the other hand, we highlighted the benefits of the Machine Translation and analysed the 2018 Language Industry Report, which specified the global translation industry very closely.

In the sixth chapter, we tried to explain the terms Post-editing and Pre-editing. Both of these translation methods are getting more and more to the forefront through progress in the Machine Translation development. We have found that the Post-editing and Pre-editing should become a part of the corpora of future translators and language experts. We have provided several reasons why Pre-editing should take place every time you use the Machine Translation. Thanks to the high-quality Pre-editing, we can easily reduce the error rate of the Machine Translation and thus improve the quality of the output text.

In the seventh chapter, we have specified the methods and objectives of our practical part. We have created 7 hypotheses, thanks to which we have been able to analyse our questionnaire and characterise the translation industry in Slovakia. Later, we specified the English translations of questions in our questionnaire. The questionnaire contained 18 questions and was filled out by translators of various ages, experiences, language combinations and preferences.

In the eighth chapter we have characterised the results and outputs of our questionnaire and analysed the individual hypotheses in more detail. The results of our questionnaire were presented in the "Discussion" and "Future recommendations" section.
We have found that the translation industry has undergone significant changes in recent years, which will be gradually reflected in Slovakia as well. Translators in Slovakia are familiar with the advantages and disadvantages of the Machine Translation. However, many of them still prefer their competencies and language abilities. Thanks to our questionnaire, we have proved that the Machine Translation is surely used by translation entities in Slovakia. However, the questionnaire also showed the fact that the trend of using the Machine Translation in Slovakia is different from that prevailing in the world. In our opinion, in the near future, the Machine Translation will be in the foreground in Slovakia and the demand for the Machine Translation will grow constantly. An important fact is that

the role of the translator in the translation process is still irreplaceable and in the vast majority of machine translations it will never replace the services of a studied and experienced language expert. We assume that, in the current situation and in the development of the Machine Translation, for some texts solely Post-editing will be required because the quality of the Machine Translation will be absolutely sufficient.

Resumé

Diplomová práca s názvom Use of machine translation tools in commercial translation in Slovakia sa skladá z úvodu, ôsmich kapitol, záveru a resumé.

Strojový preklad sa v posledných rokoch stal jedným z najdôležitejších aspektov prekladateľského priemyslu na celom svete. Vďaka technologickým pokrokom sa strojový preklad stal neoddeliteľnou súčasťou našich životov. Jeho dôležitosť narastá ešte väčšmi práve vďaka medzinárodnej globalizácii a rastúcej závislosti štátov jeden na druhom. V našej diplomovej práci sme sa pokúsili definovať frekvenciu využitia nástrojov strojového prekladu v komerčnom preklade na Slovensku.

Konkrétne sme sa pokúsili určiť okrem frekvencie využitia strojového prekladu aj to, ako prekladateľské subjekty na Slovensku vnímajú pojem strojový preklad a ako k nemu pristupujú. Prekladateľské agentúry na Slovensku vyžadujú od prekladateľov čím ďalej tým viac znalosť post-editácie a pre-editácie textu určeného pre strojový preklad. Prekladatelia preto musia čeliť znižovaniu ceny ich práce a niektorí dokonca zvažujú odchod z tohto priemyslu.

V prvej kapitole sme sa sústredili na charakteristiku a vznik pojmu strojový preklad. História a vznik strojového prekladu sa datuje do 60-tych rokov minulého storočia. V roku 1954 sa uskutočnila prvá verejná demonštrácia strojového prekladu. Tento experiment bol uskutočnený vďaka spolupráci medzi spoločnosťou IBM a univerzitou Georgetown University. Tento úspešný pokus znamenal nový začiatok pre prekladateľský priemysel. Veľké množstvo firiem a vlád chcelo finančne podporiť rozvoj strojového prekladu, pretože dúfali, že tak môžu veľmi jednoducho a efektívne nahradiť ľudský aspekt. Vďaka silnému rozmachu v oblasti počítačových technológií sa mnohí odborníci pokúsili dostať do stavu plnej automatizácie prekladov. Po desiatich rokoch výskumov zmizol prvotný entuziazmus a pojem strojový preklad takmer úplne zapadol prachom.

Ďalej sme sa v prvej kapitole zamerali na charakteristiku samotného strojového prekladu a bližšie sme si priblížili začiatky strojového prekladu na Slovensku. Pojem strojový preklad predstavuje rôzne počítačové systémy, ktoré sú schopné prekladu zdrojového textu v jednom jazyku do východiskového textu v inom jazyku. Rovnako ako technológie sa do popredia dostáva čím ďalej tým viac aj internet. Poskytuje nám množstvo informácii, ktoré sú však častokrát v inom jazyku a niekedy dokonca v jazyku, ktorému nedokážeme porozumieť. Človek jednoducho nedokáže ovládať všetky jazyky sveta a práve

z toho titulu vznikli interkultúrni sprostredkovatelia a jazykoví experti, ktorí sa nazývajú prekladatelia.

V Československu sa vznik a príchod strojového prekladu datuje do 50. rokov minulého storočia. Prvé pokusy prebiehali na Karlovej Univerzite v Prahe a získali si pozornosť primárne vedeckej komunity. Dôležitým faktom je, že v Československu prevládali pri pohľade na možnosť úplnej automatizácie procesu prekladu zmiešané názory. Bol to značný rozdiel oproti vnímaniu strojového prekladu vo svete, kde pri strojovom preklade prevažovali veľmi optimistické názory.

V druhej kapitole sme hovorili o procese a metódach strojového prekladu. Môžeme povedať, že existujú dva základné prístupy v strojovom preklade. Tieto prístupy k strojovému prekladu sú:

- Pravidlový strojový preklad (RBMT)
- Korpusový strojový preklad (CBMT)

Každý z týchto prístupov má svoje vlastné špecifiká a charakteristiky. Konkrétne sme hovorili aj o tom, ako jednotlivé procesy prebiehajú. Najkomplexnejším prístupom k strojovému prekladu je tzv. hybridný preklad, ktorý je relatívne mladý a kombinuje viaceré prístupy k strojovému prekladu. Hybridný strojový preklad prináša kombináciu najčastejšie pravidlového strojového prekladu a následne štatistického strojového prekladu alebo strojového prekladu založeného na príkladoch. Takýto typ strojového prekladu kombinuje to najlepšie z viacerých druhov strojového prekladu a výsledný preložený text je vďaka tomu kvalitnejší. Aj keď takto preložený text stále obsahuj chyby, je pre recipienta pochopiteľný a významovo korektný.

V tretej kapitole sme sa venovali nástrojom strojového prekladu. Bližšie sme si špecifikovali Google Prekladač a prekladový nástroj SYSTRAN. Charakterizovali sme ich z nášho uhla pohľadu. Bližšie sme si priblížili ich históriu a aj to, ako oba prekladače fungujú. Pre Google Prekladač sme sa rozhodli z toho dôvodu, že je to jeden z najdostupnejších a najvyužívanejších nástrojov pre strojový preklad. Do roku 2016 používal Google Prekladač štatistický strojový preklad a radil sa medzi priemerné strojové prekladače, čo sa kvality týka. V roku 2016 bola do Google Prekladača implementovaná neurónová sieť, ktorá jeho kvality pozdvihla na novú úroveň a vďaka tejto skutočnosti sa Google Prekladač postupne stáva jedným z najlepších voľne dostupných prekladačov na trhu.

Prekladač SYSTRAN patrí medzi jedny z najstarších nástrojov strojového prekladu, ktorý sa vyvíja dodnes. Tento prekladač je založený na kombinácii pravidlového strojového

prekladu a štatistického strojového prekladu. Jeho nevýhodou je, že nie je dostupný pre všetky jazykové kombinácie sveta, ale len pre tie, ktoré sa používajú najčastejšie.

V štvrtej kapitole sme sa venovali najnovšej metóde strojového prekladu a to konkrétne neurónovému strojovému prekladu. Ako sme už spomenuli v druhej kapitole, neurónový strojový preklad môžeme zaradiť medzi korpusové strojové preklady. Tento prístup k strojovému prekladu sa momentálne radí medzi tie najlepšie. Neurónový strojový preklad bol predstavený v roku 2013. Z nášho uhla pohľadu je neurónový preklad veľmi komplexná a obrovská neurónová sieť založená na umelej inteligencii a učení sa. Faktom je, že aj tento typ strojového prekladu má svoje nevýhody, no na druhej strane je jeho miera chybovosti oveľa nižšia, ako je to pri iných prístupoch strojového prekladu. Konkrétnejšie sme si porovnali neurónový strojový preklad so štatistickým strojovým prekladom. Na rozdiel od štatistického strojového prekladu sa neurónový preklad pokúša pochopiť vetu ako jeden zmyslový celok a nerozdeľuje vety na jednotlivé subkomponenty ako je to pri štatistickom strojovom preklade.

V piatej kapitole sme demonštrovali všeobecné chyby strojového prekladu z viacerých hľadísk. Konkrétne to bolo z hľadiska bežnej komunikácie, kontextu, polysémie, synoným, homoným, idiómov and dynamického vývoja jazyka. Každý z týchto aspektov predstavuje úskalie strojového prekladu, na ktoré sa vývojári budú musieť sústrediť a je veľmi ťažké predpokladať, či sa im to vôbec podarí niekedy prekonať. Jazyk ako taký je živý organizmus, ktorý sa neustále vyvíja. Na druhej strane sme vyzdvihli výhody strojového prekladu z hľadiska učenia sa a toho, že sa neurónový preklad dokáže tomuto vývoju prispôsobiť. Medzi najväčšie výhody strojového prekladu môžeme jednoznačne zaradiť schopnosť strojového prekladu prekladať medzi prakticky akýmikoľvek dvoma jazykmi. Ďalšou výhodou je aspekt efektivity pre prekladateľa. Vďaka strojovému prekladu môže prekladateľ uspokojiť väčšie množstvo klientov s porovnateľnou kvalitou výstupného textu.

Neskôr sme analyzovali 2018 Language Industry Report, ktorý sa venoval situácii prekladateľského trhu vo svete. Výsledky tohto prieskumu jednoznačne dokazujú fakt, že sa strojový preklad pomaly ale isto stáva neoddeliteľnou súčasťou či už životov obyčajných ľudí, ktorý strojový preklad využívajú pre preklad elektronickej komunikácie, alebo profesionálov, ktorí využívajú strojový preklad pre zvýšenie svojej efektivity a uspokojenie väčšieho množstva klientov.

V šiestej kapitole sme sa pokúsili vysvetliť pojmy post-editácia a pre-editácia. Oba tieto pojmy sa vďaka progresu vo vývoji strojového prekladu dostávajú do popredia čím ďalej tým viac. Zistili sme, že by sa post-editácia a pre-editácia mali stať súčasťou korpusov

vo vzdelávaní budúcich prekladateľov a jazykových expertov. Pri príprave budúcich jazykových expertov je dôležité, aby boli pripravení na čo najväčšie množstvo požiadaviek klientov pre úspešnú akreditáciu vo svojej obore.

Pri post-editácii rozlišujeme pojmy ako ľahká post-editácia a úplná post-editácia. Ľahká post-editácia znamená len minimálnu úpravu preloženého textu. K ľahkej post-editácii dochádza vtedy, keď klient potrebuje preklad urgentne a v krátkom čase. Pri úplnej post-editácii dosahuje preložený text najvyššiu kvalitu a klient nevie určiť, či bol text preložený prekladateľom alebo strojovým prekladom. Uviedli sme dôvody, prečo by pri každom použití strojového prekladu mala prebehnúť aj pre-editácia. Vďaka kvalitnej pre-editácii môžeme jednoducho znížiť chybovosť strojového prekladu a tým zvýšiť kvalitu výstupného textu. Rozdiel medzi post-editáciou a pre-editáciou je ten, že pri post-editácii prekladateľ upravuje už preložený text strojovým prekladom. Pri pre-editácii sa prekladateľ snaží upraviť zdrojový text do čo možno najjednoduchšej formy pre následný strojový preklad. Strojový preklad vysokej kvality môže zvýšiť efektivitu prekladateľa, ale následná post-editácia môže byť v niektorých prípadoch veľmi frustrujúca. Pri oboch postupoch je dôležité zachovať pôvodnú myšlienku autora a rovnako aj štýl písania. Neskôr sme uviedli príklad pre-editácie na konkrétnom texte, na ktorom sme demonštrovali dôvody, prečo je pre-editácia textu strojového prekladu skutočne dôležitá.

V siedmej kapitole sme špecifikovali metódy a ciele našej praktickej časti. Žijeme v spoločnosti, kde môžeme stretnúť kohokoľvek z rôznych kútov sveta a keďže nie je možné, aby sme ovládali všetky jazyky sveta, spoliehame sa na niekoho alebo niečo, čo nám umožní komunikovať v cudzom jazyku. Rozhodujeme sa buď pre prekladateľa alebo, čím ďalej tým viac, pre strojový preklad. Pre zistenie konkrétnych údajov sme sa rozhodli vytvoriť dotazník, ktorý obsahoval 18 otázok a zameriaval sa na viaceré aspekty prekladateľského priemyslu na Slovensku. Vytvorili sme 7 hypotéz, vďaka ktorým sme dokázali analyzovať náš dotazník a charakterizovať tak prekladateľský trh na Slovensku. Neskôr sme uvideli aj anglické preklady otázok v našom dotazníku. Dotazník bol vyplnený prekladateľmi rôznych vekových kategórii, skúseností a preferencii.

V ôsmej kapitole sme charakterizovali výsledky a zistenia nášho dotazníku a podrobne sme analyzovali jednotlivé hypotézy. Výsledky nášho dotazníku sme uviedli v časti Discussion a Future recommendations. Znenia všetkých hypotéz sú uvedené nižšie:

- 1. hypotéza: Priemerný prekladateľ sa prekladateľským službám venuje už viac ako 10 rokov.

- 2. hypotéza: Priemerný prekladateľ pracuje na menej ako 10 prekladateľských projektoch mesačne.
- 3. hypotéza: Najčastejšie prekladaným druhom textu je právnický text.
- 4. hypotéza: Angličtina a slovenčina sú najfrekventovanejšie východiskové a cieľové jazyky na Slovensku.
- 5. hypotéza: Pre prekladateľa je efektívnejšie využiť strojový preklad textu, ktorý následne post-edituje.
- 6. hypotéza: Google Prekladač je najčastejšie využívaný nástroj strojového prekladu na Slovensku.
- 7. hypotéza: Neurónový strojový preklad je pre prekladateľa kvalitatívne lepšie ako iné typy strojového prekladu.

Po zodpovedaní všetkých hypotéz sme boli schopní zhodnotiť splnenie cieľa našej diplomovej práce.

Táto práca dospieva k zisteniu, že prekladateľský priemysel prešiel za posledné roky signifikantnými zmenami, ktoré sa postupne budú prejavovať aj na Slovensku. Prekladatelia na Slovensku poznajú výhody aj nevýhody strojového prekladu. Veľká časť z nich však stále uprednostňuje svoje jazykové kompetencie a schopnosti pred využitím strojového prekladu. Vďaka nášmu dotazníku sme dokázali, že je strojový preklad prekladateľskými entitami na Slovensku využívaný v o niečo nižšej miere ako je to vo svete. Z dotazníka však vyplynul aj fakt, že je trend využitia strojového prekladu na Slovenku iný ako ten, ktorý prevláda vo svete. Podľa nášho názoru sa blízkej budúcnosti dostane strojový preklad do popredia aj na Slovensku a bude žiadaný stála viac a viac. Aj keď sme si jasne demonštrovali početné výhody strojového prekladu, stále ešte nie je dostatočne vyvinutý, aby dokázal vo väčšej miere nahradiť odbornosť a skúsenosti jazykového experta. Strojový preklad sa v posledných rokoch dostal do fázy, kedy je pri jeho vhodnom použití možné dosiahnuť vyššiu efektivitu pre prekladateľa. Dôležitým faktom je, že je úloha prekladateľa v procese prekladu stále nezameniteľná a vo veľkej väčšine strojový preklad nikdy nenahradí služby študovaného a skúseného jazykového experta.

Skúseným prekladateľom odporúčame venovať strojovému prekladu zvýšenú pozornosť. Táto technológia, ktorá sa stala neoddeliteľnou súčasťou každodenného života, dokáže byť veľkým pomocníkom. Zistili sme však aj, že sa veľká väčšina skúsených prekladateľov radšej spolieha na svoje jazykové zručnosti. Vďaka strojovému prekladu môžu zvýšiť svoju efektivitu a uspokojiť tak viac klientov.

Na druhej strane tu máme veľké množstvo prekladateľov, ktorí ešte nie sú tak dlho súčasťou prekladateľského priemyslu. Ich veľkou výhodou je to, že sa narodili do obdobia plného technológii, a preto sa s pojmom strojový preklad veľmi rýchlo oboznámili. Pre veľkú časť takýchto prekladateľov sú strojový preklad a počítačom podporovaný preklad veľmi zaujímavé. Vďaka novým technológiám sa nástroje pre strojový preklad vyvíjajú nesmiernou rýchlosťou. Odporúčame preto novým a menej skúseným prekladateľom, aby sa problematike strojového prekladu aktívne venovali, pretože sa takýto typ prekladu stane podporným nástrojom pri veľkej časti prekladov.

Môžeme skonštatovať, že prekladateľské subjekty na Slovensku nástroje strojového prekladu využívajú, ale je to v menšej miere ako je to pri prekladateľských subjektoch vo svete.

9. References

9.1. Book sources

ABSOLON, J. – MUNKOVÁ, D. – WELNITZOVÁ, K. 2018. *Machine Translation: Translation of the future?* Ružomberok : Verbum, 2018. 78 p. ISBN: 978-80-87800-45-4

BHATTACHARYYA, P. 2015. *Machine Translation.* Boca Raton : CRC Press, 2015. 217 p. ISBN: 978-1-4389-9719-5

CHESTERMAN, A. – WAGNER, E. 2014. *Can Theory Help Translators? : A Dialog Between the Ivory Tower and the Wordface.* New York : Routledge, 2014. 142 p. ISBN: 978-1-900650-49-6

HORWOOD, E. 1986. *Machine Translation. past, present, future.* Sydney : Halsted Press, 1986. 382 p. ISBN: 0-470-20313-7

HUTCHINS, J. 2000. *Early years in Machine Translation. Memoirs and biographies of pioneers.* Amsterdam : John Benjamins Publishing Company, 2000. 386 p. ISBN: 90-272-4586

KRINGS, H. 2001. *Repairing Texts : Empirical Investigations of Machine Translation Post-editing Processes.* Kent : The Kent State University Press. 2001. 627 p. ISBN: 0-87338-671-X

MUNKOVÁ, D. 2013. *Prístupy k strojovému prekladu*. Nitra : Univerzita Konštatntína Filozofa v Nitre, 2013. 115 p. ISBN: 978-80-558-0450-7.

MUNKOVÁ, D. – VAŇKO, J. et al. 2017. *Myliť sa je ľudské (ale aj strojové).* Nitra : Univerzita Konštantína Filozofa v Nitre, 2017. 260 p. ISBN: 978-80-558-1255-7

O'BRIEN, S. et al. 2014. *Post-editing of Machine Translation : Processes and Applications.* Cambridge : Cambridge Scholars Publishing, 2014. 335 s. ISBN: 1-4438-5476-X.

OCH, F. J. 2002. *Statistical Machine Translation: From Single-Word Models to Alignment Templates* : PhD thesis. Aachen : . Aachen University, 2002. 144 p.

ORHAN, Z. 1998. *Confidence Factor Assignment To Translation Templates* : Master thesis. Ankara : Bilkent University. 1998. 94 p.

SHAH, K. 2012. *Model adaptation techniques in Machine Translation* : PhD thesis. Le Mans : Le Mans University. 2012. 100p.

SVOBODA,T. 2012. *Kapitoly z překladatelské praxe*. Praha : Univerzita Karlova v Praze, 2012. 206 p. ISBN 978-80-7308-407-3

XUAN, W. H. – Li, W. – TANG, Y. G. 2012. *An Advanced Review of Hybrid Machine Translation* : International Workshop on Information and Electronics Engineering. Heilongjiang Province. 2012. 3022 p.

9.2. Internet sources

ANJU – KUMAR, M. 2014. *Malayalam To English Machine Translation : An EBMT System*. [Online]. IOSR Journal of Engineering. ISSN 2250-3021. 2014. [visited 2019-01-08].
Retrieved from
https://pdfs.semanticscholar.org/7d37/5d1e83d5332a0f0812fd93e796cf4784e340.pdf?_ga=2.69782608.1941690016.1547381342-1454397792.1547381342

ARTETXE, M. – CHO, K. 2017. Unsupervised Neural Machine Translation. *Conference Paper at the International Conference on Learning 2018*. [Online]. 2017. [visited 2019-01-12]. Retrieved from https://arxiv.org/pdf/1710.11041.pdf

BAHDANAU, D. – CHO, K. – BENGIO, Y. 2014. *Neural Machine Translation by Jointly Learning to Align and Translate*. [Online]. 2016. [visited 2019-01-11]. Retrieved from https://arxiv.org/abs/1409.0473

CADWELL, P. – HUOVÁ, K. 2016. *A Comparative Study of Post-editing Guidelines*. [Online]. 2016. [Online]. 2016. [visited 2019-01-13]. Retrieved from http://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/4_2_23_Hu.pdf

CHÉRAGUI, M. A. 2012. *Theoretical overview of Machine Translation.* [Online]. 2012. [visited 2019-01-07]. Retrieved from http://ceur-ws.org/Vol-867/Paper17.pdf

EAMT. 2008. *What is Machine Translation?* [Online]. 2016. [visited 2018-12-28]. Retrieved from http://www.eamt.org/mt.php

EUATC. 2018. *2018 European Language Survey. Expectations and Concerns of European Language Industry.* [Online]. 2018. [visited 2019-03-03]. Retrieved from https://www.euatc.org/images/2018_Language_Industry_Survey_Report.pdf

GREEN, L. 2015. *Everything you ever wanted to know about Google Translate, and finally got the chance to ask.* [Online]. 2016. [visited 2019-01-11]. Retrieved from https://www.taus.net/think-tank/articles/everything-you-ever-wanted-to-know-about-google-translate-and-finally-got-the-chance-to-ask

HAJIČ, J – HAJIČOVÁ, E – ROSENT, A. 1992. *Machine Translation research in Czechoslovakia.* [Online]. 1992. [visited 2019-01-06]. Retrieved from https://www.erudit.org/fr/revues/meta/1992-v37-n4-meta338/002996ar.pdf

HUTCHINS, J. 1993. *Latest developments in Machine Translation technology: beginning a new era in MT research*. [Online]. 1993. [visited 2018-12-27]. Retrieved from http://www.hutchinsweb.me.uk/MTS-1993.pdf

HUTCHINS, J. 2005. *The history of Machine Translation in a nutshell*. [Online]. 2016. [visited 2018-12-28]. Retrieved from http://www.hutchinsweb.me.uk/Nutshell-2005.pdf

HUTCHINS, J. – SOMERS, H. 1992. *An introduction to Machine Translation.* [Online]. 1992. [visited 2019-01-03]. Retrieved from http://www.hutchinsweb.me.uk/IntroMT-TOC.htm

JULIA, S. 2011. *Do You Know the Difference Between Pre-Editing and Post-Editing*

KASTBERG, P. 2012. *Machine Translation Tools – Tools of the Translator´s Trade.* [Online]. 2012. [visited 2019-01-02]. Retrieved from https://tidsskrift.dk/claw/article/view/7238/6172

MANNING, C., & SOCHER, R. (2017). *Natural Language Processing with Deep Learning.* Stanford : Stanford University. [Online]. 2017. [visited 2019-01-10]. Retrieved from https://web.stanford.edu/class/cs224n/lectures/cs224n-2017-lecture10.pdf

O'Brien, SCH. 2010. *MT Post-editing Guidelines.* [Online]. 2010. [visited 2019-01-3]. Retrieved from https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines

SINDHU, D. V. – SAGAR, B. M. 2017. *Dictionary Based Machine Translation from Kannada to Telugu*. [Online]. [visited 2019-01-08]. Retrieved from http://iopscience.iop.org/article/10.1088/1757-899X/225/1/012182/pdf

TUROVSKY, B. 2016. *Ten years of Google Translate.* [Online]. 2016. [visited 2019-01-11]. Retrieved from https://blog.google/products/translate/ten-years-of-google-translate/

UNKNOWN, 2014. *Light and Full MT Post-Editing Explained*. [Online]. 2014. [visited 2019-01-14]. Retrieved from https://www.rws.com/insights/rws-moravia-blog/bid-353532-Light-and-Full-MT-Post-Editing-Explained/

UNKOWN, 2019. *Pre-editing and post-editing (MT)*. [Online]. 2019. [visited 2019-01-15]. Retrieved from https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta/

WU, Y. – SCH, M. – CH, Z. et. al. 2016. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. [Online]. 2016. [visited 2019-01-09]. Retrieved from https://arxiv.org/abs/1609.08144v1

Annex

Questionnaire

Miera využitia strojového prekladu na Slovensku prekladateľmi na voľnej nohe

Dobrý deň. Som študentom 5. ročníka Fakulty aplikovaných jazykov na Ekonomickej univerzite v Bratislave. V rámci svojej diplomovej práce sa zameriavam na mieru využitia strojového prekladu v komerčnom preklade na Slovensku. Dotazník má okrem iného priblížiť frekvenciu, druh a spôsob využitia strojového prekladu v prekladovom procese profesionálneho prekladateľa/profesionálnej prekladateľky.

Dotazník Vám zaberie približne 10 minút a vopred Vám ďakujem za jeho kompletné vyplnenie. V prípade otázok, nejasností alebo pripomienok ma, prosím, neváhajte kontaktovať prostredníctvom e-mailovej adresy hanusekjakub@gmail.com.

Ďakujem.

* Povinné

1. Pohlavie *
   - Muž
   - Žena

2. Ako dlho sa profesionálne venujete prekladu? (v rokoch) *
   - 1 – 5
   - 6 – 10
   - 11 – 15
   - 16 – 20
   - 20 a viac

3. Približne na akom množstve projektov pracujete v rámci jedného kalendárneho mesiaca? (uveďte číslo) *
    - Vaša odpoveď

4. Približne na akom množstve projektov pracujete v rámci jedného kalendárneho roka? (uveďte číslo) *
    - Vaša odpoveď

5. Akej odbornej oblasti sa pri svojej práci venujete? *
    - Ekonómia
    - Právo
    - Medicína
    - IT
    - Literatúra (rozprávky, beletria a i.)
    - Publicistika
    - Iné:

6. Akému druhu textov sa pri svojej práci primárne venujete? *
    - Zmluvy
    - Návody (manuál, užívateľská príručka)
    - Správy (články, noviny a i.)
    - Iné:

7. Z ktorých východiskových jazykov prekladáte najčastejšie? (viacero možností) *
    - Angličtina
    - Bulharčina
    - Čeština
    - Čínština
    - Dánčina
    - Estónčina
    - Fínčina
    - Francúzština
    - Gréčtina

- Hebrejčina
- Holandčina
- Chorvátčina
- Japončina
- Kórejčina
- Litovčina
- Lotyština
- Maďarčina
- Maltčina
- Nemčina
- Poľština
- Portugalčina
- Rumunčina
- Ruština
- Slovenčina
- Slovinčina
- Španielčina
- Švédčina
- Taliančina
- Turečtina
- Iné:

8. DO ktorých cieľových jazykov prekladáte najčastejšie? (viacero možností) *
    - Arabčina
    - Bulharčina
    - Angličtina
    - Čínština
    - Dánčina
    - Estónčina
    - Fínčina
    - Francúzština
    - Gréčtina

- Hebrejčina
- Holandčina
- Chorvátčina
- Japončina
- Kórejčina
- Litovčina
- Lotyština
- Maďarčina
- Maltčina
- Nemčina
- Poľština
- Portugalčina
- Rumunčina
- Ruština
- Slovenčina
- Slovinčina
- Španielčina
- Švédčina
- Taliančina
- Turečtina
- Iné:

9. Využívate pri svojej práci strojový preklad (MT)? *
    - Áno
    - Nie (pokračujte otázkou č. 12)

10. Ak áno, ako často ho pri prekladateľskej činnosti využívate?
    - Pri každom preklade
    - Pri viac ako polovici prekladov
    - Pri menej ako polovici prekladov

11. Ak áno, ktoré z voľne dostupných prekladačov využívate najčastejšie? (viacero možností)
    - Google Translate
    - Microsoft Bing Translation
    - Yandex
    - Baidu Translate
    - WorldLingo
    - Bing Translator
    - ImTranslator
    - iTranslate
    - Iné:

12. Ak áno, ktoré z komerčných prekladačov využívate najčastejšie?
    - Microsoft Translator Hub
    - Kantan MT
    - Omniscien
    - SDL BeGlobal
    - Yandex API
    - Slate Desktop
    - Systran
    - Babylon
    - Iné:

13. Ak áno, ako hodnotíte kvalitu strojového prekladu pri preklade Vami zvolených jazykov? (1 = najlepšie, 5 = najhoršie)
    - 1
    - 2
    - 3
    - 4
    - 5

14. Ak nie, z akého dôvodu ho nevyužívate? (niekoľko dôvodov)
    - Vaša odpoveď

15. Ktorý spôsob prekladu je podľa Vás z časového hľadiska pre prekladateľa najefektívnejší? *
    - MT (Machine Translation)
    - CAT Tool (Computer Assisted Translation Tool)
    - PEMT (Post Editing Machine Translation)
    - Iné:

16. Stretli ste sa pri svojej práci s pojmom NT (neurónový preklad)? *
    - Áno
    - Nie

17. Využívate pri svojej práci softvér podporujúci neurónový preklad?
    - Áno
    - Nie

18. Ak áno, ako ho hodnotíte? (1 = najlepšie, 5 = najhoršie)
    - 1
    - 2
    - 3
    - 4
    - 5