IRSTI 16.31.02

B.Y. Mussagaliyeva[1], E.A. Akhapov[2]
[1,2]Al-Farabi Kazakh National University, Almaty, Kazakhstan
(E-mail: [1]i3ota@mail.ru, [2]erlan.ahapov@kaznu.kz)

Methods for evaluating the effectiveness of machine translation systems

Abstract. In the era when the Internet has been developed and the boundaries of international communication have been overcome, the need for the process of translation from one language to another is gradually increasing, since most of the information on the Internet is presented in foreign languages. To get access to information, both ordinary people and some specialists in their activities turn to machine translation systems for help. Unfortunately, the quality of machine translation has not yet reached the desired high level. Its quality depends on many factors. Scientists and translators conduct numerous studies and evaluate the work of programs in order to further improve the quality of machine translation programs. The purpose of this article is to familiarize readers with the methods for evaluating the performance of machine translation systems, and the opinions of experts on the operation of machine translation programs. The methods for evaluating machine translation described in the article were considered both theoretically and practically.

Keywords: translation, machine translation, translation systems, evaluation of translation, methods for evaluating translation.

DOI : https://doi org/10.32523/2616-678X-2020-130-1-112-124
Received: 21.01.20 / **Final version:** 24.02.20 / **Approved for publication:** 30.03.20

Introduction. We live in an era of information technology. The rapid development of new information technologies proves that the role of computer technologies in the global information space is constantly growing. Nowadays access to new information has become faster and easier. Its main part comes to us in foreign languages, which makes the work of translators more demanded. Unfortunately, people cannot always provide a quality translation in a short time. Generally, good work requires more time, and by that time the information may already be irrelevant. Therefore, more and more people are resorting to machine translation services. At the initial stage of its development, machine translation was not popular and it was believed that it had no future. However, now, machine translation systems are used not only by ordinary people, but also by specialists from various fields to optimize their work. Nevertheless, the quality of translations provided by machine translation systems still leaves much to be desired. Usually, it depends on the source and target languages, the theme and style of the original text, and the degree of familiarization with the theme of the reader. To improve the work of such systems and the quality of translations, it is necessary to evaluate the effectiveness of machine translation systems. The results of these evaluations help developers upgrade systems and provide the market with ahigh-quality product. Studies conducted to evaluate machine translation are important not only for translation studies and linguistics. The development of such systems can help in various areas of human activity.

Purpose of the paper. The purpose of this paper is to familiarize the reader with methods for evaluating the effectiveness of machine translation systems offered by various specialists and scientists.The paper describes these methods and their features. Also, the described methods have been applied in practice for more detailed analysis of the translation results.

Methodology. As methods in this article were used descriptive analysis in the theoretical part, and analysis and comparison methods in the practical part.

Results. Today, the number of people using different translation services in their activities to translate any texts from one language to another is increasing rapidly. Furthermore, with the hope that in the future, automated translation programs can satisfy the demands of people for translating from one language into another, countries spend millions of dollars to develop computer software that will be widely used to automate the translation process [1]. Along with the development of machine translation systems, there is a need to develop reliable methods for evaluating machine translation quality. The quality of translation depends on several objective and subjective factors [2]. It is primarily determined by the quality of the original text or the spoken language, as well as the professionalism of the interpreter and his willingness to carry out that specific translation act. Scientists have long been dealing with the problems of evaluating the effectiveness of machine translation systems (MTS – hereinafter this abbreviation will be used). Different approaches are offered before and are offering now: from statistical to the way of taking account of the user's opinion. The methods of assessment of the effectiveness of the current MTS can be divided into two general categories: manual evaluation of machine translation (MT – hereinafter this abbreviation will be used) and automatic evaluation of MT.The manual evaluation method is based on the evaluation of the quality of machine translation by certain criteria and a certain scale conducted by man. The manual evaluation of the quality of machine translation is very expensive, subjective and time-consuming [2]. From the very early stage of theMTS, almost all methods of evaluating machine translation, which were carried out by different specialists and scientists, fall into this manual evaluation method. Despite the diversity of methods presented, there is currently no general methodology for evaluating the effectiveness and potency of MTS. There are also opinions on the impossibility of comparing different machine translation systems due to inequality in the initial stages of such programs. How we can compare two systems, one of which is developed by a whole group of linguists, programmers with significant financial support from sponsors, and the other is developed by enthusiasts of their field? Scientists claim that to make a clear comparison, significant investments are needed to carry out identical system configurations and develop common criteria.

Marchuk Y.N. said that there is still no single criterion for assessing the effectiveness of the NSR and suggests considering the cost of the systems and evaluating its work for a certain period [3].

There are also theoretical aspects of this service, besides the practical ways of evaluating translation carried out through machine translation systems based on empirical, intuitive and professional experience and expertise. In the translation theory, the adequacy and efficiency of translation are limited to categories of semantic completeness and accuracy, which are usually supplemented with stylistic equivalents, including, in particular, the principle of correspondence of the text of the translation with the stylistic norms of the target language. Evaluation of the quality of the translation, mainly based on these parameters.According to the author, there are different types of translation adequacy: semantic-stylistic, functional and desiderative (the type of expression that expresses desire). Besides, due to the communicative purpose of the original text, the adequacy of the translation may be varied into variational, initiative, informative and selective types [2].

Novikov A. I. proposes a comparison of the original and translation texts based on the denotative (informative) sentence structure [4]. In this way, a denaturation structure is creating for each of the texts. The nodes of such a structure are concepts, and the relationships between them are predicates, that represent the relationship between nodes. The comparison is based on determining the proximity or distance between the denotative structures.

Many scientists, with a traditional approach to evaluating translation quality, offer clarity and adequacy as the main criteria for translation efficiency. Some of them noted, that for demonstrating the effectiveness of MTS performance it is necessary to test them on representative arrays of

information. One of the first examples of evaluating the effectiveness of MTS performance was the well-known ALPAC report (machine translation assessment report of Automatic Language Processing Advisory Committee, 1966) [5, 125 p.].

The clarity criteria mean how understandable the text of the translation is when it is read without reference to the original. There are various ways to evaluate the translation of the text made by the machine.It is offering to evaluate the clarity of the translation using a four-point scale, five-point scale [6], and a ten-point scale used to prepare the ALPAC report.Let's take a look at the five-point scalepresented by Makoto Nagao:

1. The meaning of the sentence is clear and no questions arise. Grammar, usage of words, and style correspond to the general structure of the text and do not require post-editing.

2. The meaning of the sentence is clear, but there are big problems with grammar, usage of words, and style.

3. The general meaning of the sentence is understandable, but the meaning of some of its parts is doubtful due to the incorrect grammatical structure, word-usage, and stylistic mistakes. Requires reference to the original.

4. The sentence has a large number of grammatical, word-usage and stylistic mistakes. The meaning of the sentence can hardly be understood after careful study.

5. The meaning of the sentence is incomprehensible. A similar classification, in the reverse order, is used to evaluate translation quality by Christopher Hogan and Robert Frederking [7]:

5. Perfect;

4. One - two mistakes, but otherwise good;

3. A few mistakes, but the meaning can be understood;

2. Some parts are translated correctly, but it is difficult to understand the meaning;

1. It is completely incomprehensible.

Some researchers, considering the concept of generalizing the scale, suggest reducing theevaluating scale on a tree-point scale:

1. Good (5 in the previous classification);

2. Acceptable (4, 3 in the previous classification);

3. Unacceptable (2, 1 in the previous classification).

Nevertheless, it is difficult to assume the unity of opinion in this field, as the ALPAC report uses a ten-point scale to make a historical decision.

However, notwithstanding the diversity of classifications, none of them can be regarded as objective. The degree of understanding of the text by a recipient often depends on the individual, consequently, and subjective factors (level of knowledge, level of familiarization with the topic in the text, etc.).

The concept of adequacy implies a mandatory reference to the original text in order to find out how accurately the translation corresponds to the original source. Such experiments are usually carried out by specialists who have a good command of both the source language and the output language in which the translation is actually performed. The adequacy criterion serves to confirm the correct transfer of the meaning of the original since there are cases when the recipient perfectly understands the content of the text, but this content does not fully correspond to the content side of the source text. For example, let's take a look at the seven-point scale of adequacy evaluating presented by Nagao [6]:

1. The meaning of the sentence in the source language (SL) fully corresponds to the content of the output sentence. There is no need for transformations; the native speaker fully understands the meaning of the output sentence;

2. The meaning of the sentence in the SL corresponds to the content of the output sentence. The native speaker understands the meaning of the output sentence, but some transformation is necessary;

3. The meaning of the sentence in the SL is correctly conveyed in the output sentence. Word order changes are required;

4. Despite the correct transmission of the general meaning of the source sentence in the output sentence, there are problems with the coordination of times, coordination between the parts of the sentence, and correct use of adverbs. Double use of nouns is possible;

5. The meaning of the proposal in the SL is not quite correctly conveyed in the output sentence. Some expressions are missed. There are problems with the coordination and correct use of the main and subordinate sentences, or parts of the sentence;

6. The meaning of the proposal in the SL is incorrectly conveyed in the output sentence;

7. The meaning of the output sentence does not fully correspond to the content of the source sentence. The structure of the output sentence does not match the structure of the normal sentence, subject and predicate are absent.

The adequacy criterion, along with the clarity criterion, are one of the most important elements of evaluating the quality and effectiveness of a translation. Unfortunately, currently there is no methodology that allows to make adequate and objective analysis of translated texts. Formalization and automation of this process requires large time and financial support, and seems impossible at the current level of technological development due to the difficulty of representing extra-linguistic knowledge in computer systems. Man-made assessment is somewhat subjective and may vary quite widely depending on the personality of the researcher.

There are various principles for assessing the current MTS along with the clarity and adequacy criteria and the size of editorial adjustments.

The «black box» principle is used to study many commercial machine translation systems by independent experts [5, 256-257 p.], with the assumption that the internal system structure of MT and its type will be based on the translations made by this system.Since many of the MTSs have product properties, the developers of the system make every effort to ensure that the principles of the functioning of the system and the translation algorithms constitute a commercial secret. In this case, the identification of the basic principles of the system's functioning is based solely on the results of translations obtained experimentally and their subsequent analysis.

In contrast to the black box principle, the so-called «glass box» transparency principle system is used [5, p.258]. Based on this principle, the performance of the system is assessed by its creators and developers. With this analysis, we can create step-by-step analysis and synthesis, determine which module or algorithm is not working properly. This principle is used to correct and fix the system by the developers themselves.

In addition, the principle of using the test arrays of texts can also be used. For this purpose, both real texts and artificially created texts are used to verify the accuracy of the translation of a particular language phenomenon. For five years, the METAL MT test system was tested on a representative array of texts, the total volume of which was about 1000 pages. Thus, for example, Margaret King and Falkedal offer a combined use of test and real text arrays in order to test the functioning of the system not only in conditions predetermined for it, but also in unpredictable conditions of real text, which is almost impossible to simulate [2].

In evaluating the effectiveness of MTS, the purpose of such a study and its direct participants play an important role. According to Margaret King, despite the above-mentioned methods of assessing the quality of modern MTS, the goal of such studies remains fundamental, since not the quality of the translation as a whole, but it's acceptable results for specific, narrow areas and tasks is subject to a real assessment.

In this regard, it is desirable to identify prospective participants and customers in the field of machine translation. The following groups are divided according to the graduation received by M. King, Trujillo, Hutchins, Somers and other researchers [5, 235-236 p.]:

1. Researchers;
2. Research sponsors;
3. Developers;
4. Buyers;
5. Translators;
6. End-users of translations.

Researchers explore the main issues of the MT and explore the possibility of their creation. The essence of the research is to identify new phenomena in the work of systems, to create prototypes of research, to verify them and to modify existing MT models.

The purpose of the research conducted by the sponsors is to identify factors that determine either in favor of continuing to finance the project, or, conversely, in favor of terminating the investment due to its lack of prospects or high costs. Sometimes not only the level of financing for a particular project, but also for a whole scientific area depends on the results of such research. An example of a negative impact on research in the field of machine translation is the results of the research published in the famous ALPAC report.

The developers of the MTS actively collaborate with the creators of the systems. This category of professionals is engaged in the practical improvement of the system by transforming the research prototype into a realistic system that has the properties of a particular product and is suitable for commercial distribution. The fundamental factor of this research is the pragmatic aspect of the functioning of the program, simplicity of interface, protected work of algorithms and the acceptable quality of translations. Developers of MTS are also involved in the development of the theoretical model in the use of existing MT systems. Quite often, in order to improve the quality of translations, these specialists conduct research aimed at improving the efficiency of the system.

The category of buyers includes a large group, starting from companies and translation services to end-users and individuals. This category of people widely uses in assessing the effectiveness of the MTS along with linguistic factors and extralinguistic criteria.These criteriainclude the cost of the system, the costs of its operation and staff training, the possibility and cost of upgrades, the effectiveness of customer support services, and the competitiveness of system developers. This category of users evaluates the MT system from a pragmatic point of view based not only on linguistic criteria of the functioning of the system but also on the extralinguistic, non-linguistic factors of its operation.

Unlike the buyers, the translators are interested in the linguistic side of the MTS functioning. Considering the quality of translation, such specialists primarily evaluate linguistic support, user interface and ways to timely replenish the system. Due to their specialization, translators can help assess and improve the linguistic component of machine translation.

The end-users of the translation can be considered as buyers, translators, and anyone who uses translations in some way in their activities. Considering the breadth «end-user» and the inability of its full disclosure in the framework of this work, we would like to consider the end user as a specialist who does not know the language of the original text. Evaluation of the translation quality by such users depends on several factors, namely: the purpose of referring to the text of the translation, whether the user is a specialist in the field described in the translation, and others. The purpose of referring to the translation may be a signal reading to get the main meaning of the document. In signal reading, the minimum quality requirements are imposed on the translation - it should be generally understood and reflect the most general meaning of the original text. For more detailed acquaintance with the text of the translation, it must correspond to certain criteria, namely, to convey the terminology most fully and correctly, the translation of which should be carried out in accordance with the latest changes in the vocabulary of both languages. The grammatical structure of the output sentence should also convey the basic structure of the input sentence. This factor does not play a decisive role when an expert refers to the text since it is possible to eliminate stylistic mistakes by deep knowledge in the subject area under consideration.

MT systems, in addition to the linguistic component, also consist of certain software. In this regard, special standards have been developed and created. Within the framework of the EAGLES / TEMAA project, a research group has been established to evaluate the effectiveness of modern machine translation systems [5, 260 p.]. To determine the quality of software for MT systems, this research group developed the standard ISO 9126. To do this, six main characteristics of software quality were identified: functionality, reliability, practicality, efficiency, openness, and portability. These criteria can also apply to all other types of software.

White J.S. says that any native speaker can estimate the translated sentence in their language as good or bad, acceptable or unacceptable as soon as they read it. In this case, they do not have to be able to think deeply, to analyze sentences, or to have a lot of linguistic knowledge [8]. However, in order to evaluate the quality of the translated text, it is important to consider both the original text and the translated text. White indicates two parameters in evaluating MT, that are considered in the manual assessment of MT: equivalence and accuracy. It follows that among people trust in automatic programs of MT assessment is very low.

As we can see, different scientists and specialists can offer different evaluation criteria and standards, but most of them can be identical and correspond to certain system settings. Another example, Callison-Burch considers that we can use the rating system to estimate the translated text [9, p.140]. That is, the appraiser must evaluate the quality of the translation in accordance with a number of criteria based on a rating for evaluating the results of various translation systems and put them in one range (from (1) The worst to (5) The best).Vilar and Stymne in their work proposed a method of mistakes analysis [10]. Mistakes analysis is the identification and classification of mistakes in the translated text.

In recent years, many translation companies and specialists have been developing automatic programs that evaluate the results of the MTS. Now they are only going through testing stages. Researchers believe that in the next five to ten years these programs will be widely used in various fields, since work on their development is being carried out intensively. Assessment metrics that are considered standard are loaded into such programs, for example, LEU, METEOR, BLUE, NIST, LEPOR [11]. Based on these metrics, an analysis is carried out and a translation estimate is provided. Programs for automatic MT assessment can open the following possibilities [2]:

• The decision to publish a translation with or without post-editing;

• Notification of the reader about the quality of the translation;

• Find sentences and phrases that require professional translator service rather than automatic edits;

• To choose the best version of translation from the presented ones.

A variety of methods and approaches for assessing the quality of MT reflects the lack of a single standard for determining ongoing research in this area and the effectiveness of existing systems. A detailed study of all methods for assessing modern MTS requires special study outside the scope of this work. Basically,MT assessment methods can be classified as «black box» and «glass box». For some reasons, many specialists use the black box method. In the evaluation of the MTS based on the principle of «black box», sometimes it is impossible to interpret the translation mistake, without considering or studying systemic algorithms. It should be notedthat translation mistakes are closely related to each other, and some mistakes can often lead to new inaccuracies. In this regard, it is necessary to apply a universal approach to the issue of assessing the performance ofMTS. Only a systematic description of the program allows the correct sorting of MTS systems and their mistakes.

For the purpose of practical consideration of the above-mentioned assessment methods, we used passages from the Kazakhstan news portal named tengrinews.kz as examples. An online translator Google Translate was used as a translation tool, and the texts were translated from the Kazakhlanguage into the Japanese language. Evaluation scales presented by Makoto Nagao,

Hoganand Frederking were used during the evaluation of translated texts.As mentioned above, assessing the effectiveness of MTS by participants directly related to the operation and use of MTS can help improve the productiveness of MT systems. Therefore, let's consider what estimates or judgments can be given by each of the categories of participants.

Researchers, as mentioned above, are engaged in the development of the main problems of MT and possible ways to solve them. Often, researchers themselves create a potential problem to consider and study in more detail possible bugs and difficulties in the operation of the system. If the system is working properly, then development is proceeding correctly and any large-scale changes in the code or device of the system are not necessary at the moment. And if the system often gives errors and does not present the desired result, then we should review the code, the structure of the program, or completely redo the system from the beginning, that can be energy and resource-consuming, and may lead to termination of work on this program. Thus, we decided to provide the following assessment scale of the MTS by researchers: 1 - during the operation of the system, no errors or problems were detected; 2 - large errors were not detected in the system operation, but fixing minor faults is required; 3 - the system is working satisfactorily, further development is required; 4 - the system is malfunctioning, fixing the code and structure is required; 5 - the system does not work.

The research sponsors, as the name implies, decide whether there are prospects and benefits to continue financial support or is it better to abandon this project and start a new one. As already mentioned, sometimes their decision can affect the whole scientific industry, so their assessment of the effectiveness of the MTS can be crucial for the development of such systems. We propose the following scale for assessing the effectiveness of the MTS by research sponsors: 1 - the results of the MTS work to justify the costs, financial support should be continued; 2 - the results of the MTS are satisfactory, it may be worth continuing to finance the research; 3 - the results of the MTS are below expected, perhaps funding should be stopped; 4 - the results of the MTS are unsatisfactory, funding should be discontinued.

Developers, in the course of their practice of refining the system, can immediately assess the operation of the MTS and therefore it is possible to determine whether the system needs modifications, how bug-resistant it is, whether its interface seems practical, and what quality shows translations in general. The final product that users and customers receive largely depends on the developers. Thus, they have an important role in creating machine translation systems. The evaluating scale of developers on the effectiveness of MTS can be represented as follows: 1 - the quality of the translations of MTS is good, the interface is handy and the system algorithm works without interference; 2 - the quality of translations is acceptable, the interface is of an average level of quality and convenience, the algorithm of the system may require improvements; 3 - the quality of translations is unsatisfactory, the interface is complicated and the system requires improvements at the code level.

We suppose that buyers and end-users of the translation can be combined into one category because both of them mean directly a group of users, or individual users of machine translation systems. They give an assessment of the work and effectiveness of MTS mainly based on their needs and requirements, and the sphere of activity of the group or individual users is also important. As mentioned above, the main criteria for evaluating by users and customers can be the cost of the system, the cost of its operation, training costs, the cost of upgrades, the competitiveness of the system, the adequacy and comprehensibility of the translations provided, whether the system can be adapted for translations in the field necessary for the user, how accurately the translation of terminology is carried out, and others. We offer the following evaluating scale by users: 1 – the translation is adequate and fully understandable, the cost and expenses are acceptable, easy to use; 2 - the translation conveys the main idea, but requires improvements, the costs are acceptable; 3 - the translation partially conveys the main idea, but justifies the cost and expenses;4 - translation is

partially understandable, not convenient to use; 5 - the translation is completely incomprehensible and cannot be regarded as adequate, is not convenient to use and does not justify the costs. The following category of translators, with their evaluation of the work of MTS, can help developers and researchers in solving the main problems that arise when the system is being used. They can also quickly identify inaccuracies and inconsistencies between the texts of the original and the translation, thus setting the right direction for further modifications. As mentioned above, translators can assist in evaluating and finalizing the linguistic component of the system. We offer the following scale for evaluating the effectiveness of MTS by translators: 1 - the translation is adequate and understandable, correctly conveys the stylistic and semantic features of the original text, grammatically correct, there is an update of the system on time; 2 - the translation is clear and adequate, but there are inconsistencies in the stylistic features of both original and translation texts, grammatically correct, there is an update of the system on time; 3 - the translation is partially understandable, there are mistakes and inconsistencies in the style and semantic features of the original and translation texts, has some grammatical mistakes, a partial system update is also observed, post-editing is required; 4 - the translation is almost incomprehensible, there are mistakes and inconsistencies in the style and semantic features of the original and translation texts, has grammatical mistakes, post-editing is required, there is no system update on time; 5 - the translation is completely incomprehensible, there are a lot of semantic and stylistic mistakes and inconsistencies, has grammatical mistakes, post-editing is required, there is no system update on time.

As it was written before, we used texts from the Kazakhstan news portal to overlook the methods described in this paper and assess the effectiveness of MTS. We randomly selected 10 texts of various subjects, which were translated by Google online web service, better known as Google online translator. This web service is designed to automatically translate part of a text or web page into another language. The translation was made from the Kazakh language into the Japanese language. We conducted an analysis of the translation texts and gave assessments of MTS for each of the evaluating scales described above. Translation evaluation results are presented in the table 1.

Table 1. Translation evaluation results

Text number	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
Makoto Nagao's evaluation scale	3	2	3	3	3	4	1	1	3	4
Christopher Hogan and Robert Frederking's evaluation scale	2	4	3	2	2	2	4	4	3	2
On an abbreviated three-point scale	3	2	2	2	3	3	1	1	2	3
Makoto Nagao's adequacy assessment	5	2	2	5	5	5	2	2	4	6
Researchers evaluation	3	2	2	3	3	3	2	2	3	4
Research sponsors evaluation	2	2	1	3	3	3	1	1	2	3
Developers evaluation	2	2	2	2	3	2	1	1	2	3
Evaluation of buyers and end-users of translations	4	2	2	4	4	4	1	2	4	4/5
Translators evaluation	3	3	2	3	4	3	2	2	3	4

As we can see from the table above, at the moment, Kazakh-Japanese machine translation is only on average, and in most cases at a low level. There are many inaccuracies and semantic distortions. It follows that improvements are still required, as well as the replenishment of the vocabulary of systems. But this may not be enough for an accurate translation, as there are cultural and linguistic features of both languages and certain rules for the use of words and styles. We carried out a detailed analysis of text translations, and determined what mistakes and inaccuracies were encountered in these examples. Let's take a look at some of the mistakes.

The translation of the first example has several semantic mistakes: the athlete «has risen» in the world ranking – has been translated as «獲得しました» (kakutokushimashita) – won; «rise 64 steps» in Japanese became «64 倍になりました» (64baini narimashita) - has increased 64 times. To some extent, these translations can be understood, but they distort the main semantic flow. There are problems with the comprehensibility of the text of the translation and it becomes necessary to refer to the text of the original. The first mistake can be assessed as "a small inaccuracy" since in general the phrases "rose in the rating" and "won" in this context can be considered similar in meaning. But, the translation of the phrase "64 steps" presented as "64 times" is rude and distorts the full meaning of the whole sentence. The machine translation provided for this text cannot be equivalent to the original text.

In the text of the translation, the abbreviation of the names of the researches is deciphered as follows: in the first one it is completely inconsistent with the original and translated incorrectly, the second abbreviation is missing, and only the third is translated correctly but into Japanese. We consider that since the abbreviations were given in the English version in the original text, we should also leave it in English in the translation text and provide an explanation or version in Japanese in brackets (as it was in the original text). The translation of the phrase «Minister of Education and Science» did not consider the difference in state structure between the two countries and there was no «one» certain translation. At the beginning of the text, the phrase is translated as "教育科学大臣" (kyouikukagakudaijin), which literally translates as Minister of Education and Science. This is the correct translation. And the second time the same phrase is translated as "文部科学大臣" (monbukagakudaijin), which means «Minister of Education, Culture, Sports, Science and Technology», and refers only to Japan. Most likely, the reader will understand the meaning of the translation text, but there may be confusion, because the Japanese reader will perceive the phrase "文部科学大臣" as the minister of his country.

The general meaning of the 3rd text's translation is understandable. There are several lexical and semantic mistakes. For example, word «Kazakhstani» is translated as "カザフスタンの品種" (kazafusutannohinshu) – sort Kazakhstan, the meaning can be understood, but it is lexically incorrect; "watched" became "watch over", usually we don't use this word about concert; "performance" translated as "コンサート番号" (konsaatobangou) – concert number, here word "number" stands for digital symbol; "Gakku Dauysy", which is the name of the concert, must remain untranslated, as it refers to the proper name, but in translation we got "Gakku Dauysy" («dauysy» means «voice» in Kazakh). Perhaps this discrepancy arose due to the fact that part of the name was written in Latin letters, and part in Kazakh. The system cannot correctly understand and translate this phrase in the context of a sentence. But, for some reasons, if you take "Gakku Dauysy" separately and translate it, we will get the correct translation.

In the 4th example general meaning is understandable, but some words and phrases are missing in the text of the translation, 2nd and 3rd sentences are combined into one. In spelling the name of the president, Latin is used first, then the Japanese alphabet for foreign words katakana, but there is no one certain method, and this can confuse the reader. "4th meeting of speakers of the parliament of the countries of Eurasia" was translated incorrectly – "ユーラシア経済共同体の議会議長の第４回会議" (4th meeting of the Chairman of the

Eurasian Economic Community) – this can be regarded as a gross mistake distorting the original meaning of the original text and the translation of the next sentences, some grammatical mistakes are observed, but the general meaning is clear. The meaning of the last sentence is conveyed incorrectly, there are grammatical mistakes, it is impossible to interpret the meaning correctly without referring to the original text, since the main semantic points of the sentence were mixed up: if the original text says that the countries signed an agreement when participating in the meeting, then the translation text indicates that the agreement was only signed, but it's hard to understand who participated in the meeting and whether it relates to the agreement.

The beginning of the translation of the 5thtext is completely understandable. In the 2nd sentence, «ol» which means "she" in this context, was changed to "Tengrinews.kz", which completely distorts the main meaning. Some parts of the original text are missing. 3 and 4 sentences were combined and part of the meaning was conveyed incorrectly: due to the fact that "大量死" (tairyoushi) was transmitted in the 3rd sentence, the following discrepancy is observed - "your empty words destroyed my dreams, my childhood and mass death".The word "she" again translated incorrectly, in the next part of the text it became "he". but this cannot be a system error, since in the Kazakh language there are no separate words for he and she. So, the system just chose the general version. This kind of mistakes makes the post-editing necessary, because the computer cannot recognize them. The general meaning of the text is understandable and with a little editing of grammar and vocabulary, it can be used as an alternative translation.

In the next text, the first sentence is complex grammatically and stylistically, it is difficult for the system to recognize which of nouns belong to the secondary and main verb. The word order is wrong;therefore, the meaning is completely incomprehensible. There is no verb at the end of the translated sentence, which further complicates the understanding of the meaning. The translation is not understandable. Post-editing is required. The last sentence is conveyed more correctly, the meaning is clear, there are some grammatical inaccuracies.

In the translation of the 7th text, the translation is accurate and understandable. There are minor grammatical mistakes, but they do not prevent the reader from understanding the general meaning. Perhaps there is an incorrect use of certain words, in particular specific terminology. From which it follows that it may be difficult for an ordinary reader who is not familiar with the subject of the original text to understand the translation well enough.

Translation of the 8th text can be an example of a good machine translation. But it is worth noting that in the original text there are no complex sentences, there are also no phraseological units or complex phrases. The structure of the text is not complicated and easy to understand even for people who do not have special knowledge. 1stsentence is accurate and understandable, but 2nd is missing. The translation conveys the general meaning and is done quite well. In the last sentence there are some lexical inaccuracies that can complicate the understanding of the text by a native speaker of the target language. The translation is good, but post-editing may be necessary. Unlike the previous example, the translation of the 9th text cannot boast of good quality. There are many inaccuracies and mistakes. We believe that the degree of comprehensibility of the original text can be estimated as medium complexity, due to the presence of complex sentences and special terms. With a detailed analysis, we get the following. The name of the prime minister is not translated correctly. Translation is inaccurate and partially incomprehensible; the words are confused and some are translated incorrectly - "10 percent" was translated as "October 10th". Also, it contains semantic inaccuracies and difficulties for understanding. There are inaccuracies in the translation of ministerial posts. Short sentences are translated properly and convey the meaning accurately, clearly. There are grammatical and lexical inaccuracies; post-editing is required. Also, the text of the translation will be difficult to understand without reference to the text of the original.

The translation of the 10th example is also unsatisfactory. As in the previous examples, for the translation system, there are problems with complex sentences and phrases of the original text, and the presence of special terminology, apparently an unfamiliar to the system for this particular pair of languages. In a detailed analysis, we encounter the following mistakes and inaccuracies in the translation. The word "com" was used inappropriately twice, the word coma was translated incorrectly, it must be "昏睡" or "コマ" in Japanese, but the translation system could not recognize this word correctly. There is no verb related to the main subject. The next sentence does not correspond at all to the original: in translation we see the following - "they were searched for a woman's body", when there should be "artificially supported the woman's body". The meaning is distorted also in the subsequent sentence: «the birth of a girl is good» should be - «the girl was born healthy.» The phrase "an amazing event" is translated as "amaze". And in the last sentence, there is again distortion of the meaning of the original text, and some mistakes in word translation: for example, the phrase "she came to her senses" is translated as "she was in trouble".

Let's summarize all of the above. In almost all translation examples, there is a discrepancy between style and lexical structure with the original text. Incorrect translation of words also took place in all examples, in some 1-2 words in the text, in some 1-2 words in the sentence.

In all examples, there are cases of the absence of a word or a whole phrase, even a sentence. For example, there are sentences in which the main verb is missing, there is no main noun, or the words were mixed up. The names of events that are untranslatable, in many examples were distorted, partially translated, and sometimes there were 2 different translations in one text. The same is observed with the names of people, in the text of the translation they can be written first in Latin letters, and then again, but in Japanese alphabet, or vice versa, which indicates the absence of standards for translating proper names in MT system for a given pair of languages. There are also examples where in the text of translation 2-3 sentences were combined into one, and because of the incorrect word order the meaning of the originaltext was completely distorted.In 9 out of 10 cases, post-editing of the translation text is required.

Conclusion. At the moment, there is still no generally accepted methodology for evaluating machine translation systems. Different developers and researchers may suggest different methods and assessment scales. Nevertheless, direct users remain as mainin the assessment of translations of MTS, who also contribute to the development of such systems: demand today solves a lot.In our case, demand also has a great influence on Kazakh-Japanese machine translation. We can see the difference between the Kazakh-Japanese and Kazakh-Russian MTS only by comparing the final results, and this difference will clearly not be in favor of the Kazakh-Japanese machine translation. It is also necessary to develop a generally accepted system for evaluating machine translation. This could help in correcting errors and bugs of the MTS, as it would be clearly visible in which direction to move.

Thus, we can conclude that at the moment, machine translation from the Kazakh language into the Japanese language cannot be completely satisfactory. In some cases, we can evaluate such a translation as good, but it is still far from "excellent". Machine translation systems for this particular pair of languages require improvements and modifications.

References

1 Ртищева Н.Г.Оценка качества работы систем машинного перевода // Решетневские чтения. – 2017. – Т. 2. – С. 704-705.

2 Улиткин И.А.Автоматическая оценка качества машинного перевода научно-технического текста //Вестник МГОУ.Сер.Лингвистика. – 2016. – Т. 4.– С.174-182.

3 Марчук Ю.Н. Проблемы компьютерной лингвистики; Модель «текст-текст» и переводные соответствия в теории машинного перевода //Сборник научных статей. – Минск, 1997. – С. 21-29.

4 Новиков А.И.Применение денотатной структуры текста для перевода научно-технической литературы //Психолингвистические аспекты грамматики. – Москва, 1979. – С. 66-75.
5 Trujillo A. Translation engines: Techniques for Machine Translation.–Springer, 1999. –P. 303.
6 Nagao M., Mori S. A New Method of N-gram Statistics for Large Number of n and Automatic Extraction of Words and Phrases from Large Text Data of Japanese //COLING 94: In Proceedings of the 15th International Conference on Computational Linguistics,Kyoto, Japan, 1994. – Kyoto, 1994. –P. 611-616.
7 Hogan C., Frederking R.E. An Evaluation of the Multi-engine MT Architecture // In Lecture Notes in Artificial Intelligence 1529. – Berlin, Heidelberg, New York: Springer-Verlag, 1998. -P. 113-123.
8 White J.S. How to evaluate machine translation // Computers and Translation: A Translator's Guide. – Philadelphia: John Benjamins, 2003. -P. 211–244.
9 Callison-Burch C., Fordyce C., Koehn P., Monz C., Schroeder J. Second Workshop on Statistical Machine Translation. – Prague, 2007. –P.158.
10 Stymne S., Ahrenberg L. On the practice of error analysis for machine translation evaluation. [Электрон.ресурс]. – 2012. – URL: http://www.mt-archive.info/LREC-2012-Stymne-2.pdf (Қаралғанкүні: 15.07.2019)
11 Boitet C., Blanchon H., Seligman M., Bellynck V. Evolution of MT with the Web // Machine Translation 25 Years On. International Conference, Cranfield, - England, 2009. – P. 13.

Б.Е. Мұсағалиева, Е.А. Ахапов
Әл-Фараби ат. Қазақ ұлттық университеті, Алматы, Қазақстан

Машиналық аударма жүйелерінің нәтижелілігін бағалау әдістері

Аңдатпа. Қазіргі интернет желісі дамыған уақытта және шекарасыз әлемдік қарым-қатынас заманында, интернеттегі ақпараттың көпшілігі шет тілдерде ұсынылғандықтан, бір тілден екінші тілге аудару процесін жүзеге асыру қажеттілігі бірте-бірте өсіп келеді. Қажетті ақпаратқа қол жеткізу үшін қарапайым адамдар да, қандай да бір маман өкілдері де өз қызмет аясында машиналық аударма жүйелерінің көмегіне жүгінеді. Өкінішке орай, машиналық аударманың сапасы әлі де қанағаттанарлық жоғары дәрежеге жете қойған жоқ. Оның сапасы көптеген факторларға байланысты болады. Ғалымдар мен аудармашы мамандар машиналық аударма бағдарламаларының жұмысын жетілдіру мақсатында көптеген зерттеулер жүргізіп, бағдарламаның жұмысын бағалайды.
Мақаланың мақсаты оқырманды машиналық аударма жүйелерінің жұмысын бағалау тәсілдерімен, мамандардың аударманы бағалау туралы ой-пікірімен таныстыру. Мақалада қарастырылған машиналық аударманы бағалау тәсілдері практика жүзінде қолданылған.

Түйін сөздер: аударма, машиналық аударма, аударма жүйелері, аударманы бағалау, аударманы бағалау әдіс-тәсілдері.

Б.Е. Мусагалиева, Е.А. Ахапов
Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

Методы оценки эффективности систем машинного перевода

Аннотация. В эпоху развитой сети Интернет и безграничной международной коммуникации, потребность в процессе перевода с одного языка на другой постепенно возрастает, поскольку большая часть информации предоставляется на иностранных языках. Чтобы получить доступ к информации, как обычные люди, так и некоторые специалисты в своей деятельности обращаются за помощью к системам машинного перевода. К сожалению, качество машинного перевода еще не достигло желаемого высокого уровня. Его качество зависит от многих факторов. Ученые и переводчики проводят многочисленные исследования и оценивают работу программ чтобы в дальнейшем улучшить качество программ машинного перевода. Целью данной статьи является ознакомление

читателей с методами оценки работы систем машинного перевода, и мнением специалистов о работе программ машинного перевода. Методы оценки машинного перевода, описанные в статье, были рассмотрены теоретически и практически.

Ключевые слова: перевод, машинный перевод, системы перевода, оценка перевода, методы оценки перевода.

Reference

1 Rtishcheva N. G. Ocenka kachestva raboty sistem mashinnogo perevoda [Evaluation of the quality of machine translation systems], Reshetnevskie chteniya [Reshetnev readings,], 2,704-705 (2017). [in Russian]
2 Ulitkin I.A. Avtomaticheskaya ocenka kachestva mashinnogo perevoda nauchno-tekhnicheskogo teksta [Automatic quality assessment of machine translation of a scientific and technical text],Vestnik MGOU. Ser. Lingvistika [Bulletin of MGOU: Linguistics], 4,174-182 (2016). [in Russian]
3 Marchuk Yu.N. Problemy komp'yuternoil ingvistiki; Model' «tekst-tekst» I perevodnye sootvetstviya v teori i mashinnogo perevoda [Problems of computer linguistics; The text-to-text model and translation correspondences in machine translation theory], Sbornik nauchnyh statei [Collection of scientific articles],Minsk, 1997. P. 21-29. [in Russian]
4 Novikov A.I. Primenenie denotatnoi struktury teksta dlya perevoda nauchno-tekhnicheskoi literatury [The use of the denotative structure of the text for the translation of scientific and technical literature], Psiholingvisticheskie aspekty grammatiki [Psycholinguistic aspects of grammar], Moscow, 66-75 (1979). [in Russian]
5 Trujillo A. Translation engines: Techniques for Machine Translation.(Springer, 1999, 303p.).
6 Nagao M., Mori S. A New Method of N-gram Statistics for Large Number of n and Automatic Extraction of Words and Phrases from Large Text Data of Japanese, COLING 94: In Proceedings of the 15th International Conference on Computational Linguistics, Kyoto, Japan, 1994. P. 611-616.
7 Hogan C. Frederking R.E. An Evaluation of the Multi-engine MT Architecture, In Lecture Notes in Artificial Intelligence 1529, Berlin, Heidelberg, New York: Springer-Verlag, 1998. P.113-123.
8 White J.S. How to evaluate machine translation, Computers and Translation: A Translator's Guide, Philadelphia: John Benjamins, 2003. P. 211–244.
9 Callison-Burch C., Fordyce C., Koehn P., Monz C., Schroeder J. Second Workshop on Statistical Machine Translation.(Prague, 2007, 158).
10 Stymne S., Ahrenberg L. On the practice of error analysis for machine translation evaluation [Machine Translation Archive] 2012. [Electronic resource]. Available at: http://www.mt-archive.info/LREC-2012-Stymne-2.pdf (Accessed:15.07.2019)
11 Boitet C., Blanchon H., Seligman M., Bellynck V. Evolution of MT with the Web, Machine Translation 25 Years On, International Conference, Cranfield, England, 2009. P. 13.

Авторлар туралы мәлімет:

Мұсағалиева Б.Е. – Шығыстану факультетінің магистранты, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.
Ахапов Е.А. – PhD, «Жапонтану» бөлімінің меңгерушісі, Шығыстану факультеті, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.
Mussagaliyeva B.Y. – Master`s student, Oriental studies faculty, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
Akhapov E.A. – Head of the Japanese Studies Department, Oriental studies faculty, Al-Farabi Kazakh National University, Almaty, Kazakhstan.