THE AUTOMATED TRANSLATION PECULIARITIES

ЛОБЫНЕВА Екатерина Ивановна

кандидат исторических наук, доцент кафедры иностранных языков для инженерных направлений Института филологии и языковой коммуникации

ORCID: 0000-0002-4053-1750

НИКОЛАЕВА Альбина Юрьевна

кандидат культурологии, доцент кафедры иностранных языков для инженерных направлений Института филологии и языковой коммуникации

ФГАОУ ВО «Сибирский федеральный университет»

г. Красноярск, Россия,

ORCID: 0000-0003-2060-6344

Automated translation main types are considered in this paper. The most popular automated translation systems working with Russian language are analyzed and compared. The principles, advantages and disadvantages of their work are stated and described. The authors suggest the possible ways of similar systems development.

Keywords: automated translation systems, language treatment, computer linguistics.

Modern world political and economical situation requires the Technical Higher School educational programs to pay attention to students' ability of using automated translation as their one of the language skills. Automated translation is the process of logical translation from one native language to others with the help of different computer programs [3].

The widespread misunderstanding is to consider automated translation and translation made by electronic resources to be one and the same thing. Automated translation is defined as fully automated software which can translate the origin content into required languages, though automated translation includes the whole translation process. That is why it contains working process of careful reading out the translated texts and edition with professional translators.

The main objectives of automated translation using:

1. Information fast gaining because automated translation lets to obtain the general information about a book, a research work or an article.

2. Communicative opportunities due to automated translation' ability to let any person communicate or write to all necessary addressees from any part of the world. Automated translation skills give the Technical Higher School graduates the chance for working in high-growing companies and increase the international market.

3. Simplifications of automated translation or interpreters work because using this they can accelerate text translation [1].

At present time there are four types of automated translation – Rule-Based Machine Translation, RBMT [3]; Statistical Machine Translation, SMT; Hybrid Translation Systems, HS; Neural Automated Translation [4].

Rule-Based Machine Translation is based on grammar rules systems which analyze the texts to be translated. The translation is carried out with the help of built in dictionaries for the language pairs and grammar comprising semantics, morphological and syntax principles of both languages. Taking into account all the data, dictionaries translate the texts consistently, sentence by sentence, for the piece of information demanded in the necessary language. The main principle of such systems work is the consideration of original and final texts structures. The grounded on grammar rules translation methods are often divided into three additional subgroups:

1. Word-by-word translation.

2. Transfer systems and interlinguistic systems.

The difficulties of this method are in its depending on lexis and it means that the effectiveness can be achieved after a long time period and the languages should be added by hand. Also it demands a lot of a translator's post edition work due to its low quality and can be used in very simple situations when only fast text understanding is preferable.

The advantages of these methods are gram-

mar and syntax accuracy, result stability and opportunity to be set to a certain subject field. But this method's disadvantages are necessity of such base creation, support, development and renovation, labor intensity and their high price.

Statistical methods apply statistical analyses while working. Their purpose is to define the interrelations between words, phrases and sentences in both languages. Bilingual texts corpus (containing a big text amount in original language and its automated translation into required one) are loaded into the system and after that it analyzes Interlingua conformity syntax structures, etc. the system is self-studied because the whole choosing of translation variant it relies on the statistic data obtained beforehand. The larger the dictionary of bilingual pair and more accurate it is the better statistical automated translation result is. Future translation quality becomes better with every newly translated text. Statistical methods differ in setting quickness and translating new approaches addition easiness. Among disadvantages there are numerous grammar mistakes and translation instability. That is why this method can be used only for base (gist) translation.

Here is the opportunity that people can translate much better than computer systems because they get the content of original texts and decode the meanings in required languages. It is rather difficult cognitive operation. To decode the whole meaning of the translated text, a translator must interpret and analyze all the text peculiarities. It is the process which demands deep knowledge in original language grammar, semantics, syntax, phrasal verbs, idioms, etc. and the native speakers' culture. A translator should have a lot of skills to decode the meaning into required languages.

There is a great automated translation problem – how to program a computer to understand the text as people do and which will create a new text in a required language coinciding with the translation made by people. Without automated translation database it can provide only general information about the translated text – so called "gist". This is enough for many purposes for example in limited and expensive time of a translator reserved for cases when full accuracy is demanded.

Hybrid translation systems include approaches written before and are the combination of Statistical translation systems and Rule-Based Machine Translation. They use translation memory and make the whole translation quality better. They are expected to give the opportunity to unite all above mentioned systems advantages. But despite of the advantages demand a translator's post edition work.

Neural Automated Translation is a method of automated translation based on neural connections in human brain. It uses artificial intelligence for learning languages and constant improvement of the knowledge. It becomes more accurate and quicker after training. Google translate was the first automated translation mechanism applying language neural treatment which can be learned after some practices.

In addition to the four above considered automated translation methods and systems we would like to analyze other ones. There are the three biggest of the automated translation systems that work with the Russian language – Abbyy Compreno, Google Translate and Yandex. Translation, each of these systems developing actively, improving and increasing the quality of translation.

Abbyy Compreno is a technology developed by the Russian company Abbyy, as well as a system of the same name created by it. For translation, the system builds a complete syntactic tree of the source text. To build it, it uses separately defined rules for each language. Moreover, Compreno is able to correctly create anaphora, ellipsis and other language constructions that are difficult to process. The translation takes into account the relationships between words in this tree, the possible position of these words in USH (Universal Sematic Hierarchy), as well as the statistics of the relationship between words. The main disadvantage of this system is its closeness, the lack of data on full-scale testing of this system, which makes it impossible to assess the quality of machine translation produced by it.

Google Translate is a system developed by Google, originally developed by Franz Josef Och. The system is based on the statistical method of translation. For the training of the system, parallel UN text bodies presented in the six official languages of the United Nations were initially used. This system does not pro-

vide direct translation for language pairs that do not include English. That means that when translating, for example, from Russian to French, the translation from Russian to English will be carried out first, and after that – from English to French. Moreover, for some languages there are more such steps. For example, texts in English are first translated into Russian, then into English, and only then turns into a text in the target language. This significantly reduces the accuracy of translation and makes the system completely unsuitable for translating a number of texts.

Yandex Translation is a system of automatic translation of words, phrases, whole texts, as well as web pages. Like Google Translate, Yandex Translation uses statistical translation. Yandex Translation consists of two parts – a translation model and a language model. The translation model is engaged in the construction of a graph containing all possible variants of the translation of the sentence. The language model selects the best translation option in terms of optimal word combinations in natural language. Although Yandex Translation has a smaller list of languages available for translation than that of Google Translate, so the translation is carried out directly, without using intermediates.

Nowadays the most common machine translation systems are based on statistical analysis. These systems are easier to create and maintain. In addition, a huge amount of data is available for training these systems, as they have at their disposal parallel text bodies, books translated into different languages, websites with the ability to choose a language and many other data sources. In contrast, rule-based systems are too time-consuming to develop and maintain. For example, it took about 15 years for Abbyy to create its Abbyy Compreno [2].

However, rule-based systems have a higher translation quality than systems based on statistics. In addition, they are more promising. It should be understood that with an increase in the number of statistical data, the translation quality of systems based on statistics increases non-linearly, and one day there will be a situation after which an increase in the number of data will no longer give at least some noticeable improvement in the translation quality. That is why the development of rule-based systems will be of greater interest. But creating systems based on rules is a necessary and promising task, but incredibly time – consuming. Dozens of linguists have been working on creating the rules for many years. Therefore, the best option would be to create a hybrid machine translation system that will have openness – thanks to which it becomes possible to replenish the rule base not only by a limited group of people, but by the entire community [5]. The hybrid principle is a combination of statistical and natural language technologies, which will allow achieving a significant improvement in the translation quality having a smaller base of rules. The possibility to work with systems using the forces of the Russian NLP community (a scientific community dealing with the problems of computer processing of natural language) has been successfully proven by the Open Corpora and NLPub projects.

The types of machine translation systems are discussed in the article; examples of such systems are provided, compared and analyzed. Rule-based systems and hybrid systems are of the greatest interest. Future researchers and developers of machine translation systems should pay attention to the creation of open hybrid machine translation systems that would be developed and supplemented by the forces of the scientific community.

REFERENCES

1. *Андреева А.Д.* Обзор систем машинного перевода / А.Д. Андреева, И.Л. Меньшиков, А.А. Мокрушин. – Текст: непосредственный // Молодой ученый. – 2013. – № 12(59). – С. 64-66. – URL:https://moluch.ru/archive/59/8581/ (дата обращения: 12.01.2023).
2. Гибридный машинный перевод: Искусственный интеллект. Системы и модели. – URL:http://www.rriai.org.ru/statisticheskiy-mashinnyiy-perevod.html (дата обращения: 12.01.2023).
3. СЛОВАРИ И СИСТЕМЫ МАШИННОГО ПЕРЕВОДА: IT LAND. – URL:http://www. itland.com.ua/products/sect.php?SECTION_ID=306 (дата обращения: 12.01.2023).
4. Статистический машинный перевод: Википедия. Свободная энциклопедия — URL:http://ru.wikipedia.org/wiki (дата обращения: 12.01.2023).

5. *Batunova I.V., Lobinova E.I., Nikolaeva A.U.* Application of effective teaching methods in foreign language classes in non-linguistic (technical) universities). Ekaterinburg 1(67, Part 4), 11–14 (2018).
6. RULE BASED MACHINE TRANSLATION SYSTEM: Natural language processing. – URL:http://language.worldofcomputing.net/machine-translation/rule-based-machine-translation.html (дата обращения: 12.01.2023).