ADVANTAGES AND DISADVANTAGES OF MACHINE TRANSLATION

Veaceslav CULAXIZ, student

CZU: 81`322.4

grottban@mail.ru

Cercetările în domeniul traducerii automate progresează în fiecare zi. Scurta istorie a traducerii automate datează aproximativ 70 de ani. În această perioadă de timp, oamenii de știință din domeniul lingvisticii și informaticii, au început să interpreteze chiar și definiția conceptului de „traducere automată” în diferite moduri (ca domeniu de cercetare, precum și ca funcție specifică a traducerii). Desigur, în timp, această direcție s-a dezvoltat, ceea ce a dus la apariția diferitelor tipuri de sisteme și abordări ale traducerii automate. Investigația acestor dificultăți au și constituit scopul prezentului articol.

The research in the field of machine translation is progressing every day. The brief history of machine translation roughly goes back about 70 years. During this specific period of time, scientists in the field of linguistics and informatics, even the very definition of the concept of "machine translation" began to be interpreted in different ways (as a field of research, as well as a specific function of translation). Of course, over time, this direction has developed, which led to the emergence of various types of systems and approaches to MT.

The first MT devices did not have extensive linguistic databases and sophisticated data analysis algorithms. They used the Direct Approach to MT, which consisted of processing the input data and directly generating the result by literally translating each vocabulary unit of a sentence or text. Some time later, the developers created a rule-based MT system, which will later be called RBMT. This system allowed for a better translation, observing (as far as possible) a number of lexical and grammatical conditions.

Significant progress in the development of MT was achieved after the emergence of a statistical translation model called SMT. The presence of a text corpus, a linguistic database, allowed the MT system to select more accurate translation options. Later, the systems began to be

combined in hybrid models, and then the developers completely implemented artificial intelligence into the MT program. As a result, a state-of-the-art MT system called NMT was developed and is still in use today.

Machine translation is capable of processing texts of various types, both literary and specialized. When translating each type, it is necessary to take into account a number of its specific features. In addition, the research repeatedly mentions stylistic (mostly, idioms), lexical-semantic (polysemy) and grammatical (verb-nominal transformations) difficulties of translation by MT systems. The above classification, a hierarchical system with five classes (Missing Words, Word Order, Incorrect Words, Unknown Words, Punctuation) and their subclasses, can help in identifying and analyzing such errors.

In general, the most numerous MT errors are those that somehow distort the meaning of the original information. Such errors are represented by the most branched class of hierarchical classification (Incorrect Words). More specifically, the most frequent errors fall into two subclasses: Wrong Lexical Choice and Incorrect Disambiguation [1].

The second subclass in terms of the number of errors, both in the translation of literary texts and specialized information, is Incorrect Form, which refers to syntactic-grammatical constructions.

Other categories of errors are extremely few in number, and, often, uncharacteristic of modern translation programs. Errors of the Unknown Words class and the Extra Words subclass occurred only once during the analysis of both cases (in general, twice). However, even a single occurrence of such errors is considered rare. They arise, as a rule, as a result of a software failure (at the stage of generating the translation result), due to an incorrect analysis of the information context, or an error in the identification of any vocabulary unit.

In general, as expected, software translation does a better job at analyzing and processing official register texts. Such texts are usually strictly structured and have many well-known parentheses that are known to translation programs like Google Translate or Yandex. MT has much more problems with literary texts. The main problems are the preservation of the style of the source material and the translation of phraseological units, which MT systems are not yet able to translate

properly. A common problem for both categories of texts is the ambiguity of vocabulary units, which are not always translated correctly.

It should also be noted that Yandex is better at analyzing the context of the input information. In many cases, thanks to the context analysis, the Yandex system was able to preserve the meaning of the original text (in particular, this concerns literary material). Google Translate showed a more accurate selection of translation equivalents for terminology and specific structures of specialized texts. Nevertheless, the accuracy and unambiguity of the selection of translation options did not always turn into an advantage. Sometimes the final result was very neutral or ambiguous, which did not match the style of the original.

Obviously, machine translation, like any emerging research area, has a number of advantages and disadvantages [2].

Advantages

- ***Speed***. Any modern activity is characterized by high rates of functioning, including work in the field of translation and interpretation. Machine translation is able to provide translation of large amounts of information, which helps to solve both work tasks and the problem of *intercultural interaction.*
- ***Availability***. Nowadays, even free translation programs are powered by a neural machine algorithm, which is one of the latest developments in the field. Anyone who has Internet access can use MT systems.
- ***Confidentiality***. A person, working directly with the program as a translator, knows for sure that information will not be disclosed by unauthorized persons.

Disadvantages

- ***Inaccuracies in translation***. In the process of analysis, the intermediate phase of translation and generation of the result, the MT system is not always able to select the correct equivalent or lexical and grammatical structure.
- ***Problems with analyzing the context of information***. The program cannot always preserve the logical coherence of the original information. Also, some expressions should be translated solely accor-

ding to the context, which is why a translation of inadequate quality is generated.

- ***Inability to perceive emotionally charged expressions***. Machine language is unable to identify the intonation of a phrase. As a consequence, the MT system makes mistakes in the translation of humor, irony, etc.
- ***Constant punctuation errors***. Punctuation systems in different languages are too different, and therefore the development of an MT algorithm for punctuation rules for each language is currently impossible [2].

Taking everything into consideration, it is important to say that MT field is progressing day by day. Thanks to the development of text corpus and artificial intelligence technologies, the quality of MT has increased drastically in recent years. The program does an excellent job in translating texts of a specialized type, with a clear structure and an official register. Yet, even at this level of quality, human post-editing is still required.

References:

1. GONCHAROV, A.A., BOONTMAN, N.V., NUREYEV, V.A. Errors in Machine Translation: Problems of Classification. In: *Systems and means of inform.*, 2019 no.3. vol. 29, [Digital Resource]; Access Date: 11.04.2021; https://www.researchgate.net/publication/340792802_machine_translation_errors_problems_of_classification
2. BELOBORODOVA, E.A., BODULEVA, A.R. The Advantages and Disadvantages of Machine Translation. In: *Innovative Science*, 2016, no.9, Russia: BSU. [Digital Resource]. Access Date: 11.04.2021. https://cyberleninka.ru/article/n/preimuschestva-i-nedostatki-mashinnogo-perevoda/viewer

Recomandat
Svetlana CORCODEL, lector, asistent univ.