MACHINE TRANSLATION AND ITS PRINCIPLES OF CLASSIFICATION

Natallia Pushyk[1]

Machine translation is an action performed on a computer that converts text from one natural language into equivalent text in another language.

Among the principles of classification of machine translation we distinguish translation with the participation of a computer in the translation process, the number of languages that the system «understands», the type of documents during the translation the system is focused on, and purely linguistic principle.

1) Translation with the participation of a computer in the translation process. The systems of machine translation are divided into automated and automatic. The automated systems of machine translation mean that the translation is performed by a machine and the term «machine-aided translation» is used for this type of translation.

---

[1] King Danylo University, Ukraine

The automated systems of machine translation indicate machine translation as such.

In automated machine translation systems, a person is a mandatory participant in the translation process. The only part of the machine participation is the performance of different routine operations, i.e. searching for words and phrases in the automatic dictionary and outputting their translation to the screen with a possible subsequent insertion into the source text. In automatic systems it is, on the contrary, the machine analyzes the input of the text and synthesizes the output, so that the translation with its inevitable lexical, grammatical and stylistic errors is generally understandable to the user and can be used as an informative document. At the same time, such systems usually have the means to edit the machine product, which is especially important during the translation of the texts that have the status of documents in the literal sense of the word, such as technical documentation supplied abroad with exported equipment [6].

2) By the number of languages that the system «understands» we mean the translation according to which we distinguish bilingual and multilingual systems. Bilingual systems translate for one language pair. If the system covers more than one language pair, it is multilingual. In the development of this point, it is advisable to take into account such divisions of the machine translation systems as «two-way systems» which translate the text in both directions, and «1-way systems» which translate the text in one direction.

3) By type of documents, the translation of which the system is focused on. In particular, some systems that work only with the titles of the documents (for example, patents), while others deal with the documents mostly with technical specifications, etc. There are also universal systems that can translate texts of different types at the same level of efficiency. Modern systems software depends only on the structure of linguistic support, but not on the specific content of the automatic dictionary, thus it is necessary to distinguish between dictionaries of different systems, as well as the practical possibility of their distribution. According to a purely linguistic principle, the classifications of the machine translation system are divided into:

a) Systems that implement direct translation (direct translation); b) Systems based on transfer (transfer approach); c) Systems with an intermediary language (Interlingua approach).

These systems are of particular interest to computational linguistics. They, according to the plan of their authors, must implement I, II, and III levels of «understanding» of the input and output texts. Moreover, it seems quite natural that the authors of this classification believe that the higher the level of «understanding» is, the higher is the quality of the translation.

a) Direct translation systems are built based on the following considerations: we have two specific languages to which this particular system is configured. The majority of the information is found in vocabulary, therefore, it is enough to translate the vocabulary of the input text correctly, for which it is necessary to remove the ambiguity that complicates this (primarily grammatical) and brings the output structure to the minimum necessary correspondence to the input syntactic structure, that the vocabulary of the source text being translated conveys the information contained in the text. These systems develop the philosophy of descriptor automatic systems. They focus on vocabulary, ignoring the «global» syntax of the sentence and relying on a minimal context. The latter is considered as a quasi-syntactic formation: the grammatical characteristics of words and their mutual arrangement are analyzed, but the syntactic, hierarchical organization of a sentence or (at least) its fragments are not. It is important to emphasize that in direct translation a sentence is both analyzed and synthesized not in the form of a syntactic formation, but as a set of linear fragments. Of course, syntactic connections and dependencies are taken into account in this linearity, but indirectly and far from in full. The essence of direct translation is in the hypothesis according to which the redundancy of the language is such that the correct translation of vocabulary compensates for grammar errors. We can also assume that the closer is the syntactic formatting of texts in two languages, the higher is the quality of translation [7].

b) The rough direct translation contrasts with the transfer approach. Its idea is if in direct translation the vocabulary often «does not work» due to disregard of the syntactic role of words, it is necessary to identify the syntactic structure of the input sentence, transform it into the structure of the translated language and only then translate it at the lexical level, that is, to substitute words into the syntactic structure of the source text.

A translation performed by a transfer system will be better than a translation performed by a direct translation system if these conditions are met:

– The system correctly recognized the syntactic structure of the text;

– The vocabulary of transfer rules is quite complete;

– The structures of the input and output texts differ so that the transformation is necessary [2, p. 597].

As we can see, the difficulties in both methods are similar: the need to predict the behavior of words in the text in advance, ie at the level of speech. In fact, in direct translation we are not interested in syntactic connections, more precisely, in their whole mass, which means that there are simply no errors that appear in non-existent connections. But, on the other hand, the transfer approach in some cases eliminates the errors of direct translation,

because it does not make direct, «subordinated» substitutions of words, and therefore does not make the correspondent mistakes.

c) According to systems with an intermediary language, one of the main prerequisites for their construction is the desire to significantly simplify the development of multilingual systems, as well as significantly improve the quality of translation. The main idea here is that, regardless of the language, the text on the stage of analysis is transformed not into a structural expression related to this language, but into a description independent of specific languages, which conveys the content of the input text. After that, the semantic representation is transformed into a text in the source language. Of course, experienced translators first understand the content of the translated text, and only then work with this content. When implementing this tool, linguists face the same problems as the developers of transferred systems, except that these problems are much more connected with the involvement of broad semantic information, which, as we have seen, is very difficult to formalize.

It is possible that different systems will translate the same text with approximately the same quality indicators, but will make different mistakes. This hypothesis needs to be tested to find new, more effective means of translation. One of such ways may be the development of hybrid systems that combine the capabilities of direct translation and transfer.

References:

1. Abney, S. (1989) A computational model of human parsing. *Journal of Psycholinguistic Research*, 18, pp. 129–144.

2. Bonnie J. Dorr. (1994) Machine translation divergences: a formal description and proposed solution. Comput. Linguist. 20, 4 (December 1994), 597–633.

3. Chandioux J., Grimaila A. (1996) 'Specialized' machine translation. In: AMTA (1996), 206–211.

4. Hutchins W.J. (1993) Latest developments in machine translation technology: beginning a new era in MT research. In: MT Summit (1993), 11–34.

5. Lee A. (1994) Controlled English with and without machine translation. Aslib Proceedings 46(5), 131–133.

6. Matsak Zh., Skorobogatova T. Mashynnyj pereklad ta jogho specyfika. [Machine translation and its specifics]. Available at: http://www.rusnauka.com/31_ONBG_2009/Philologia/54653.doc.htm (accessed 04 May 2021).

7. Novikova, O. (2018) Lexicographic resources and machine translation. *Challenges and perspectives*. Available at: http://hdl.handle.net/10347/18253 (accessed 03 May 2021).