



PROBLEMS OF MACHINE TRANSLATION

Valchuk K.I.

*National Technical University of Ukraine
"Kyiv Polytechnic Institute",
Kyiv, prospekt Pobedy, 37, tel. 044–336–7989,
e-mail: krvalchuk@gmail.com*

The idea of machine translation appeared in the 17th century. In 1629, René Descartes proposed a universal language, with equivalent ideas in different tongues sharing one symbol. The field of "machine translation" appeared in Warren Weaver's Memorandum on Translation (1949). The first researcher in the field, Yehosha Bar-Hillel, began his research at MIT (1951). A Georgetown University MT research team followed (1951) with a public demonstration of its Georgetown-IBM experiment system in 1954. MT research programs popped up in Japan and Russia (1955), and the first MT conference was held in London (1956). Researchers continued to join the field as the Association for Machine Translation and Computational Linguistics was formed in the U.S. (1962) and the National Academy of Sciences formed the Automatic Language Processing Advisory Committee (ALPAC) to study MT (1964). Real progress was much slower, however, and after the ALPAC report (1966), which found that the ten-year-long research had failed to fulfill expectations, funding was greatly reduced.

Since early 80s, when PCs have started to conquer the world, time of their work got down in a value and the machine translation finally has become economical. In addition, improvement of the programs for these and next years allowed to translate many types of text quite accurately, however, some problems of machine translation still have been staying unresolved.

Sometimes we replace term "machine" with "automatic", which does not affect the sense. Although the term "automatic translation" has different meaning: it is a program that helps a person to translate some kind of text.

Using automatic translation, we expect the following forms of interaction:

- Partly automatic translation: for example, when a translator uses computer dictionaries.

- Systems with division of labor: computer is educated to translate only phrases with a rigidly set structure (the result is not required to be corrected afterwards), but everything beyond the scheme should be worked up by a person.

Nowadays, there are two different terms in the English terminology: machine translation, MT (completely automatic translation) and machine-aided or machine-assisted-translation, MAT (automated). If there is a need to determine both, M(A)T is used.

Moreover, there are two fundamentally different ways to construct an algorithm for machine translation: rule-based and statistical-based.

The current stage of machine translation development can be described as a stage of syntactical word translation. A fundamental unit of the sense is a word, but



the definition of grammatical forms and the sequence of words in a generated text based on syntactical relationships between the words in the original text. Similarly, to the formed rules the structure of dictionaries for translations reminds a word dictionaries' structure, in which case the percentage of collocations is quite low.

In spite of the fact that the word is not the only unit of the sense in native languages, there are a small number of levels of semantic units. Among them morphemes level, word level, collocation level, sentence level, etc. What is more, the sense of the higher level units not always may be determined from the sense of units of the lower level (most meaningfully it might be traced in the morpheme and word layers relationship). Accordingly, the dictionaries for an automatic translation must contain units of various layers.

Depending on style and purpose of the text, the same word could have had different meanings. In some extent this feature is being considered in the systems of machine translation, for instance, removable dictionaries are provided sometimes even for all types of text.

Besides established rules of sentence construction, there are several unpre-scribed rules in each language called the pulchritude of a language. For example, a literal translation of a sentence "This is my pen" in Russian is "Это есть моя ручка" which is formally correct, but Russian speakers do not talk in such way. In this case, it looks like someone weird or foreigner has constructed the sentence.

In contrast, capital letters and reductions are fraught with another trick. When a word starts with a capital letter, its translation also starts with a capital letter. The effective abbreviations, read as a unit word, are frequent in English literature. A computer will also translate such abbreviations as unified words.

Sometimes it is better to perform a word-processing before translation. Thus, machine translation systems can make mistakes because of reductions ended with a point. If a word with a capital letter follows such a point, computer perceives the point as the end of the sentence. As a result, we get an erroneously parsed and incorrectly translated sentence at the output. Therefore, it is better to remove the points after reductions. There must not be any word wraps in a translatable text. Text redactor Microsoft Word can easily provide it. The texts resulting in the identification of a scanned image needs a special attention because remained errors often are being the reasons that some words left untranslated in a case of one wrong letter.

Thus, automated dictionary service should take an important place in the systems of machine translation, whose tasks include creation and management (supplementation, correct). The role of a person in the automated dictionary service is marking foreign texts in order to segregate dictionary elements and to set appropriate equivalents in the output language. A machine can also perform a text marking.

To conclude, a computer still cannot fully replaced a translator. Is there any sense to apply machine translation systems at all? The answer is absolutely there are. If a computer translates literary texts, we can get a rough version of texts, which grow to the masterpiece under the hand of a person who does not have strong language skills but is just as good as literary editor. Talking about technical text translation, in a point of using an appropriate dictionary, a quiet passable is not requiring the following treatment result may get at the output. Actually, the need of redacting a machine