

## PRELIMINARY TEXT PREPARATION AND POST-EDITING IN THE MACHINE TRANSLATION PROCESS

Incredibly rapid progress in the field of electronic hardware and computer software and attempts to receive a useable translation in the smallest amount of time have encouraged the development of *Machine Translation (MT)*, sometimes referred to as *automatic or instant translation* – translation of a text by a computer, with no human involvement. The key merit of MT is the capacity to handle very large volumes of text quickly. Due to the limitations of MT, documents will certainly contain lexical, grammar and syntax errors, though the gist of the article can be discovered for internal use.

On analyzing the scientific papers by the foreign (M. A. Cheragui, P. Koehn, M. D. Okpor) and Ukrainian (P. Asoyants, A. Mishchenko, V. Masyukevych, V. Yakovyna, A. Yankovets) scholars, information available on some MT software developers' websites and our own MT teaching practice, we **aim at** summarizing the guidelines which may be used to produce better quality MT from the English language. The **object** of research is the MT process involving preliminary text preparation (pre-editing), MT itself and post-editing stages. The **subject** of research is the ways to improve the quality of MT by changing the quality of the provided text.

Pre-editing should not be confused with post-editing. Pre-editing refers to the preparation involved prior to running a particular text through an automatic translation tool. It involves reviewing the content for basic errors, tagging certain content to be translated a certain way (or to not be translated at all), and optimizing the formatting. Post-editing is the process of “fixing” MT output to bring it closer to a human translation standard [2].

*Pre-editing of a text* can considerably simplify the job of both the machine translation systems and editors tasked with refining the raw machine translation. Such

preparation begins at the stage of composing source text. The summarized guidelines cannot solve all the problems of machine translation, but they will help to win some points in opposition of a computer and the natural language:

1. Avoid misprints and spelling mistakes. The machine translator cannot correct mistakes and recognize incorrectly spelled words (special spell-checking programs are very useful for this purpose) [1].

2. Avoid compound sentences, conjunctions (and, but, which, etc.) and homogeneous parts of sentences. Keep all the sentences short and concise. Try to use syntactic constructions with direct word order. Ideally, each sentence should convey a single logical thought.

3. Use the active voice instead of passive voice. The active voice is a direct writing style that cuts out vagueness and ambiguity. It is very difficult for MT engines to successfully translate vague phrases or those with double meanings.

4. Use infinitive verb forms instead of gerunds.

5. Use the definite article. Specify nouns using “the”.

6. Repeat nouns instead of pronouns. This improves the clarity of sentences.

7. Avoid omitting syntactic words, even if the grammar allows it [1].

8. Mind punctuation marks. Omitted or, on the contrary, redundant punctuation marks may prevent the electronic translator from correctly understanding the sentence’s syntactic structure. Paragraph marks (¶) are automatically deleted by the program, and thus the lines become one line. Therefore, it is necessary to use a period (.) at the end of the sentence [1].

9. Use diacritics correctly. As a rule, the electronic translator cannot recognize words containing the Russian letter “ë” or words with emphasis marks [1].

10. Observe the case of letters. A lowercase letter in a word may easily be converted to a capital one (for example, at the beginning of the sentence, or in the header), and it is taken into account when developing MT systems. On the contrary, a capital letter seldom becomes a lowercase one, and in most cases it relates to derivation of a new word. Besides, there are languages where capitalization changes the part of speech to which the word belongs. A striking example is the German

language, where nouns are capitalized both at the beginning and in middle of the sentence [1].

11. Use only conventional abbreviations. Even a single not translated word may prevent the electronic translator from analyzing the syntactical structure of the sentence correctly (abbreviations participate in syntactical links alongside common words). Spelling of some abbreviations often coincides with spelling of frequently used words, which may have unpleasant consequences [1].

12. Avoid using slangy expressions [1].

13. Avoid clichés/colloquial phrases. A MT engine may not convey the correct meaning of a cliché or colloquial phrase and it may not make sense to the users.

14. Write phrases that you can recycle throughout your documentation. A MT engine can recognize and accurately translate repetitive phrases.

*Post-editing* involves the refinement of a raw machine translation by an editor, who normally has special training and is experienced in handling machine-translated texts. In post-editing MT content, the process involves the following steps: 1. A sentence or segment of the MT output is read. 2. The MT output is compared to the source text. 3. Post-editor applies some form of quality judgment on each segment, based on explicit instructions. 4. Quick decisions are made on whether the translation is correct. 5. The text is improved or re-translated.

**Conclusions.** It has been proven through practice that incorporating an efficient pre-editing process, correct adaptation of MT systems, including “training” in a particular subject area, can considerably simplify post-editing of the target text. MT with post-editing can be worthwhile when dealing with large volumes of similar texts. **Further research** may be carried out in the field of customization and improvement of MT systems.

#### **References**

1. How to improve the quality of machine translation [Online resource]. – Access Mode: <http://www.online-translator.com/Help/Tips/>

2. Post-editing guidelines [Online resource]. – Access Mode: [https://kantanmt.com/documents/Post-Editing\\_Guidelines.pdf](https://kantanmt.com/documents/Post-Editing_Guidelines.pdf).