

УДК 519.6

Oleksandr Slastnikov<sup>1</sup>, Svitlana Voitenko<sup>2</sup>

<sup>1</sup>student of group CST-817 ZNTU

<sup>2</sup>senior teacher ZNTU

## **RULE-BASED MACHINE TRANSLATION TECHNOLOGY VS STATISTICAL MT**

Machine translation (MT) is automated translation. It is the process by which computer software is used to translate a text from one natural language (such as English) to another (such as Ukrainian).

To process any translation, human or automated, the meaning of a text in the original (source) language must be fully restored in the target language, i.e. the translation. While on the surface this seems straightforward, it is far more complex. Translation is not a mere word-for-word substitution. A translator must interpret and analyze all of the elements in the text and know how each word may influence another. This requires extensive expertise in grammar, syntax (sentence structure), semantics (meanings), etc., in the source and target languages, as well as familiarity with each local region.

The software parses text and creates a transitional representation from which the text in the target language is generated. She uses these complex rule sets and then transfers the grammatical structure of the source language into the target language. Users can improve the translation quality by adding their terminology into the translation process. Human and machine translation each have their share of challenges. For example, no two individual translators can produce identical translations of the same text in the same language pair, and it may take several rounds of revisions to meet customer satisfaction. But the greater challenge lies in how machine translation can produce publishable quality translations.

Statistical machine translation utilizes statistical translation models whose parameters stem from the analysis of monolingual and bilingual corpora. Building statistical translation models is a quick process, but the technology relies heavily on existing multilingual corpora. A minimum of 2 million words for a specific domain and even more for general language are required.

**Rule-based MT** provides good out-of-domain quality and is by nature predictable. Dictionary-based customization guarantees improved quality and compliance with corporate terminology. But translation results may lack the fluency readers expect/

**Statistical MT** provides good quality when large and qualified corpora are available. The translation is fluent, meaning it reads well and therefore meets user expectations. However, the translation is neither predictable nor consistent.

The differences between them are: **rule-based MT** has high performance and robustness, knows grammar rules and gives out-of domain translation quality; **statistical MT** shows good fluency and is good for catching the rule exceptions.