Ualsher Tukeev[1], Marek Miłosz[2], Diana Rakhimova[3]

QUALITY AND ECONOMIC EFFICIENCY OF MACHINE TRANSLATION METHODS

Problems of using computer methods for text translation and its economical efficiency are presented in this paper. Main methods of machine translation (MT), their characteristics and quality are introduced. The paper presents a detailed model to calculate additional losses associated with the quality of translation. Experimental method of determining the model parameters is presented. We also present the results of using the economical model to compare the efficiency of different MT systems.

Keywords: *machine translation, quality, economic efficiency.*

Уалшер Тукєєв, Марек Мілош, Діана Рахімова

ЯКІСТЬ ТА ЕКОНОМІЧНА ЕФЕКТИВНІСТЬ МЕТОДІВ МАШИННОГО ПЕРЕКЛАДУ

У статті представлено результати аналізу проблем використання комп'ютерних методів перекладу тексту і його економічної ефективності. Описано основні методи машинного перекладу (МП), їхні характеристики та якість. Розроблено детальну модель для розрахунку додаткових витрат, пов'язаних з рівнем якості перекладу. Подано також експериментальний метод визначення параметрів моделі. Узгоджено результати застосування економічних моделей для порівняння ефективності різних систем МП.

Ключові слова: *машинний переклад, якість, економічна ефективність.*

Уалшер Тукеев, Марек Милош, Диана Рахимова

КАЧЕСТВО И ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ МЕТОДОВ МАШИННОГО ПЕРЕВОДА

Проблемы использования компьютерных методов для перевода текста и их экономической эффективности представлены в этой статье. Описаны основные методы машинного перевода (МП), их характеристики и качество. В статье представлена детализированная модель для расчета дополнительных потерь, связанных с уровнем качества перевода. Экспериментальный метод определения параметров модели представлен в статье. В нем также приведены результаты применения экономических моделей для сравнения эффективности различных систем МП.

Ключевые слова: *машинный перевод, качество, экономическая эффективность.*

Introduction. Researches in the field of machine translation (MT) are carried out from the 50-s of XX century. At that time, many MT software applications (systems) were created. The software implements a specific algorithm on a text to translate into a given language pair L1 - L2, and given the direction of translation (for example: L1->L2). MT systems contain a bilingual dictionary with grammatical information, which is necessary to make the transition of words from one language to the other. In addition, they contain grammatical algorithms for the analysis of text using formal grammar (Andy & Gawronska, 2007).

[1] Prof., Department of Information Systems, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
[2] PhD, Deputy Director Institute of Computer Science, Lublin University of Technology, Poland.
[3] PhD student, Department of Information Systems, al-Farabi Kazakh National University, Almaty, Kazakhstan.

Improvement of MT quality can be achieved through the use of a man in the translation process. For example, some MT systems may translate more correctly, if the words in the text, which are names, are clearly indicated (Fiederer & O'Brien, 2009). This process is called pre-translation. The man can also choose different variants of translation, the best in a particular situation. He/she may also improve the translated text. This process is called post-translation.

Complex cognitive algorithms are used in MT. MT systems should analyse text grammatically, semantically and syntactically (Rakhimov & Zhumanov, 2011). In this analysis, the nature of a text and its authors should be taken into account. The same language can have many variations depending on the location of use (urban / rural, geographical region), social class, or cultural group (scientists vs. prisoners). The same knowledge about the target language is essential. An important problem in MT is the quality of translation. Of course, the more formal style of the document is, the more accurate the translation would be (Specia et al., 2009).

MT methods and the assessment of results' quality. There are many MT methods. The most important of them are (Callison-Burch et al., 2008) rule-based (transfer-based, interlingual and dictionary-based), statistical, example-based and hybrid. Hybrid methods usually combine the features of rule-based methods with statistics ones.

The quality of translation is related to the quality of the transfer of meaning of a text from one language to another. This quality can be evaluated using different characteristics, such as accuracy, terminology correctness, intelligibility, naturalness, linguistic correctness etc. There are many different methods to assess the quality of translation. The most important are: reverse translating, comparison with a pattern, transformation method, denotative method, surveys and error analysis on the linguistic level (vocabulary, grammar). It should be noted that the nature of natural language makes completely objective and fully formalized method of evaluation impossible, although such attempts are made (Chan & Ng, 2008; Snow et al., 2008; Klein & Manning, 2003). The most commonly used is the method of expert assessment, despite its complexity and high costs.

Expert evaluation method of MT quality. In each method of quality assessment, it is necessary to have metrics of characteristic measurement. Metrics for qualitative characteristics are usually created artificially using a point scale. This metrics, if used consistently, allows quantitative comparison quality of the evaluated systems. In assessing the quality of MT results, 3 features are the most important (Tukeyev et al., 2011):

- lexical accuracy,
- syntactic correctness,
- sense correctness (meaning).

For each single sentence, every above-mentioned characteristic is evaluated on the scale from 1 to 5 according to the criteria (Table 1). Total ratings point for a sentence is an evaluation of the translation quality. The higher the value is, the higher the quality is. A sentence can have the maximum of 15 points. To evaluate a MT system, there should be a large number of translations and their evaluation should be statistically averaged (Tukeyev et al., 2011).

Table 1. **Quality evaluation scale in the expert method**

Mark	Lexis	Syntax	Meaning (semantics)
1	All words translated incorrectly	The sentence structure is entirely incorrect	Incorrect translation
2	1-2 words translated correctly	The sentence structure is incorrect, but there are 1-2 words used grammatically correctly	Barely understandable meaning of sentences
3	About a half of words translated correctly	About a half of words are used grammatically incorrectly	General sense is understandable
4	There are 1-2 words translated incorrectly	Correct structure of sentence, but there are 1-2 words used grammatically incorrectly	The meaning of the translated text is understandable except insignificant details
5	All words are translated correctly	Correct structure of a sentence	The meaning of the translated text is correct and understandable

Source: (Tukeyev et al., 2011).

The economic model of losses caused by bad translation. Regardless the metrics and the methods for measuring and evaluating the quality of MT, it is important economically. Poor quality of translation makes the results impossible to understand in an appropriate manner and requires corrections (post-translation) by a human translator (O'Brien, 2005; Paul et al., 2007). Editorial correction has a specific economic dimension. It requires labour of a person with proper (usually high) qualifications.

During the economical analysis the MT software cost (defined as the cost of buying software licenses and the cost of its use) is relatively small and can be ignored. Modern software for MT works very quickly and is often available for free. Even if used as commercial software, the cost (due to very high speed of translation) spread over a large number of texts translated, reducing the unit costs.

Costs of employee's productivity loss, related to poor quality of text translation, can be calculated as the employer's cost of additional work made by employee-translator for correcting translation errors. So, it is possible to calculate it as (Milosz & Milosz, 2005):

$$KRU = \frac{52 * TSK * TSU}{NTR * TSP^2} KRP, \quad (1)$$

where:

KRU - average yearly cost of loss of productivity per one employee $/year;

TSK - average employee week work time with translation works, hour per week;

TSU - average employee week time loss due to translation errors, hour per week;

NTR - number of working weeks per year;

TSP - average employee week work time, hour per week;

KRP - yearly cost of employee work, $/year;

52 - number of weeks in a year.

If *KRP=1,* the formula (1) determines the percentage of loss of annual cost.

Usually, the amount of translation work is measured by standard pages. One standard page contains 1800 characters (including spaces and punctuation).

Workload of translators is measured in standard pages per day or week (Сколько, 2012). At the same time quality assessment uses a measurement of one sentence - translation unit. Conversion factor page-sentence depends on the language (the target language is usually taken) and the type of text (technical, literary, simple and complex sentences etc.).

Average employee week time loss due to translation errors may be calculated as a time for its correction:

$$TSU = PPW * TUP * \sum_{i=1}^{3}\sum_{j=1}^{5}(QLR_{ij} * ATC_{ij}), \quad (2)$$

where:

PPW - weekly load of employee in translation standard pages, pages per week;

TUP - conversion factor - average translation units per page, sentences per page;

i - type of errors - quality translation area assessment (1 - lexis, 2 - syntax, 3 - meaning (semantics));

j - number of quality assessment level ( );

QLR_{ij} - average quality level rate; it means percentage of errors of each type and level in the whole translated text; % of the errors;

ATC_{ij} - average time spent by an employee on correction of error due to its type and level; hours per error.

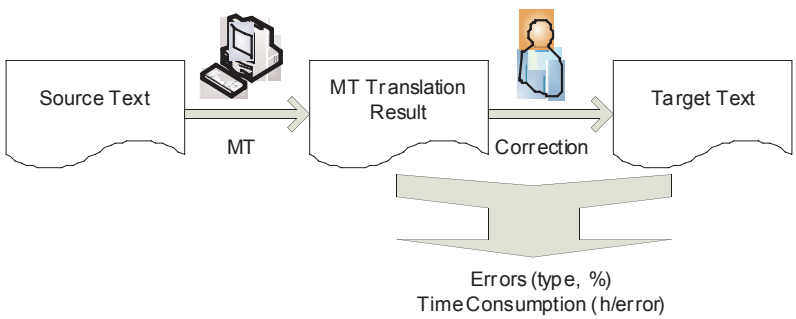
For each of groups of errors types should be:

$$\sum_{j=1}^{5} QLR_{ij} = 1, i = 1,2,3. \quad (3)$$

The economical model (formulas (1)-(3)) to calculate additional losses associated with the level of translation quality for particular firm/country and kind of translation work (in this case some parameters of model are known) can be presented as a function of computer translator quality:

$$KRU = f(QLR_{ij}) \quad (4)$$

To use the model to specific assessments, parameters of particular MT programs should be determined experimentally. Diagram of such an experiment is shown in Figure 1.



Source: own elaboration.

Figure 1. **Scheme of the experiment to determinate model parameters**

Experiment on quality assessment of different MT systems. To compare quality of different MT systems, we use the source text in Russian in the volume of a standard page (1735 characters without spaces, 279 words, 20 sentences, i.e. TUP = 20), common topic. The text was translated by 3 different MT systems into Kazakh. Two products used were Kazakh: Soylem (http://audaru.soylem.kz) and Sanasoft (http://www.sanasoft.kz/a/node/60), and one Ukrainian: Pragma6 (http://online.translate.ua/ru). Their ranges of functionality and interface are similar.

To assess the quality of the translated texts, the expert method was used. After translation of the source text with the MT software, we determine the quality parameters of 3 elements of the assessment as a percentage (the maximum number of points for 20-sentence text was 300). The translated text has been corrected by the translator, which allows determining the average time to improve each of these errors. The result of the quality assessment of 3 MT systems by expert method is shown in Table 2.

Table 2. **Results of translation quality assessment for 3 MT systems**

MT system	Total quality of translation, points	Percentage of lexical errors in the text	Percentage of syntax errors in the text	Percentage of semantic errors in the text	Average time of errors correction works, hours per error
Soylem	212	28 %	26%	34%	0.0539
Sanasoft	183	36 %	37%	37%	0.0490
Pragma6	159	45%	43%	53%	0.0536

Source: own research.

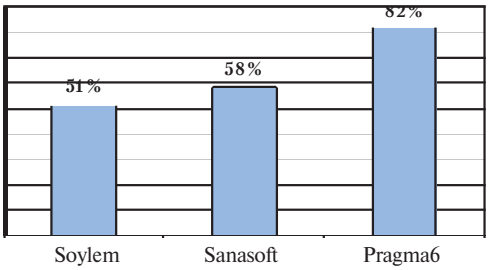
During the experiment we assessed the effort of each type of error correction made by experienced translator, and the average for each MT system (Table 2).

Experienced translator, working with the original text of medium complexity, can translate 7-8 standard pages (1800 characters) a day without losing quality. Translation efficiency increases to 12 pages per day if a text is not original to a translator (Сколько, 2012). This rate of productivity can be regarded as more or less normal. In the calculations the fact was taken into account that an employee performs the translation of full-time work (TSK=TSP) and normative indicators for Kazakhstan are: NTR=48, TSP=40, PPW=60. It was also assumed that KRP=1, which can determine the loss in % of employee time.

Figure 2 shows the results of calculations of employees' time losses to improve the text translated by MT. The best system was Soylem. Its translations require about 51% of the nominal time of an employee for corrections. This means that an employee has 49% of work time to spend on additional work. These are the employer's savings. In case of the worst MT system (Pragma6) savings from its use are very small (about 18%).

Conclusions. The economic cost of correcting errors in the translated text depends on the quality of MT. The methodology of expert assessment of quality of MT systems using 3 main parameters (lexical, syntactic, and semantic) is presented in this paper. The advantage of this methodology is that each parameter can be calculated separately, because for different types of errors different time is spent on text

correction. This property has been used to determine the parameters of the economic model, which has been developed to assess the economic losses associated with poor translation. This model can be used to determine the loss of productivity of workers and to compare the quality of MT systems.



Source: own calculation.

Figure 2. **Time lost by an employee to improve errors in the translated text**

The presented result of the experiment and calculations using the developed model allowed for comparative assessment of different MT systems, as well as the quantitative assessment of their quality in the economic dimension.

References:

Albrecht, J., Hwa, R. (2007). Regression for Sentence-Level MT Evaluation with Pseudo References, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, pp. 296-303.

Andy, W., Gawronska, B. (eds) (2007). Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation - TMI 2007, 7-9 September. Skovde, Sweden, retrieved from: http://www.mt-archive.info/TMI-2007-TOC.htm.

Babych, B., Hartley, A., Sharoff, S. (2007). Translating from under-resourced languages: comparing direct transfer against pivot translation, Proceedings of MT Summit XI, 10-14 September 2007, pp. 64-70.

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., Schroede,r J. (2008). Further Meta-Evaluation of Machine Translation, in: Proceedings of the 3rd Workshop on Statistical Machine Translation, Columbus, pp 70-106.

Chan, Y. S., Ng, H. T. (2008). MAXSIM: A Maximum Similarity Metric for Machine Translation Evaluation, in: Proceedings of ACL-08/HLT, pp. 55-62.

Сколько страниц в день может перевести переводчик?, retrieved from http://langinfo.ru/index.php?sect_id=1306

Fiederer, R., O'Brien, S. (2009). Quality and Machine Translation: A realistic objective?, Journal of Specialised Translation, No. 11, pp. 52-74.

Klein, D., Manning, Ch. D. (2003). Accurate Unlexicalized Parsing, in: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 423-430.

Milosz, M., Milosz, E. (2005). Hidden Costs of IT Skills Ignorance, Vilnius University, Information Sciencies, Vol. 34, pp. 30-34.

O'Brien, S. (2005). Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Text Translatability, Machine Translation, No. 19, Vol. 1, pp. 37-58.

Paul, M., Finch, A., Sumita, E. (2007), Reducing Human Assessment of Machine Translation Quality to Binary Classifiers, in: Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation, Skovde, pp. 154-162.

Rakhimova, D., Zhumanov, Zh. (2011). Quality Evaluation of the Machine Translation Systems, in: Varia Informatica 2011 PTI, Lublin, pp.129-146.

Snow, R., O'Connor, B., Jurafsky, D., Ng, A. (2008). Cheap and Fast - But is it Good Evaluating Non-Expert Annotations for Natural Language Tasks, in: Proceedings of EMNLP. pp. 254-263.

Specia, L., Cancedda, N., Marco Turchi, M., Cristianini, N. (2009). Estimating the Sentence-Level Quality of Machine Translation Systems, retrieved from http://mt-archive.info/EAMT-2009-Specia.pdf .

Tukeyev, U.A., Melby, A.K., Zhumanov, Zh.M. (2011). Models and algorithms of translation of the Kazakh language sentences into English language with use of link grammar and the statistical approach, in: Proceedings of the IV Congress of the Turkic World Mathematical Society, 1-3 July, Baku, 2011, p.474.

Стаття надійшла до редакції 01.01.2013.